

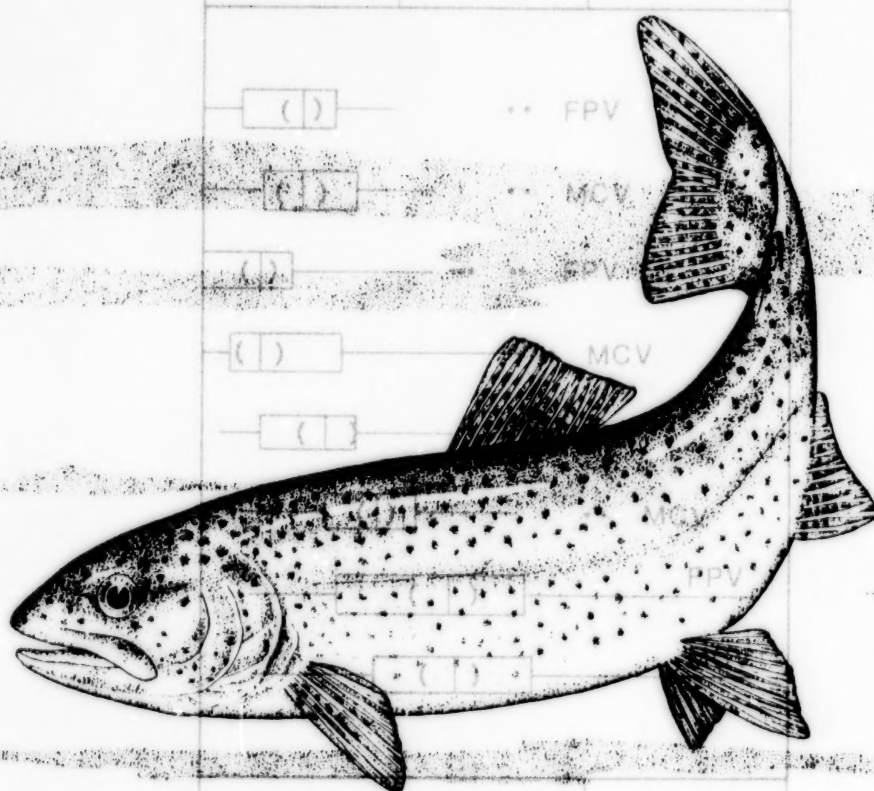
APR 14 1989

mf - 0611-R-01) (390)

I49.89/2:89(6)

BIOLOGICAL REPORT 89(6)
NOVEMBER 1988

GRAPHICAL AND STATISTICAL PROCEDURES FOR COMPARING HABITAT SUITABILITY DATA



Fish and Wildlife Service

U.S. Department of the Interior

Scf
VL



Preserve Our Natural Resources



DEPARTMENT OF THE INTERIOR
U.S. FISH AND WILDLIFE SERVICE



As the Nation's principal conservation agency, the Department of the Interior has responsibility for most of our nationally owned public lands and natural resources. This includes fostering the wisest use of our land and water resources, protecting our fish and wildlife, preserving the environmental and cultural values of our national parks and historical places, and providing for the enjoyment of life through outdoor recreation. The Department assesses our energy and mineral resources and works to assure that their development is in the best interests of all our people. The Department also has a major responsibility for American Indian reservation communities and for people who live in island territories under U.S. administration.

Biological Report 89(6)
November 1988

GRAPHICAL AND STATISTICAL PROCEDURES FOR
COMPARING HABITAT SUITABILITY DATA

by

William L. Slauson
TGS Technology, Inc.
P.O. Box 9076
Fort Collins, CO 80525-0800

Project Officer

Carl Armour
U.S. Fish and Wildlife Service
National Ecology Research Center
2627 Redwing Road
Fort Collins, CO 80526-2899

U.S. Department of the Interior
Fish and Wildlife Service
Research and Development
Washington, DC 20240

DISCLAIMER

The opinions and recommendations expressed in this report are those of the author and do not necessarily reflect the views of the U.S. Fish and Wildlife Service, nor does the mention of trade names constitute endorsement or recommendation for use by the Federal Government.

Suggested citation:

Slauson, W.L. 1988. Graphical and statistical procedures for comparing habitat suitability data. U.S. Fish Wildl. Serv. Biol. Rep. 89(6). 58 pp.

11

CONTENTS

	<u>Page</u>
FIGURES	iv
TABLES	v
ACKNOWLEDGMENTS	vi
INTRODUCTION	1
EXPLORING AND DESCRIBING CURVE DATA	3
STATISTICAL ANALYSIS OF CURVE DATA	22
The Mann-Whitney-Wilcoxon Test	25
The Kruskal-Wallis Test	27
The Kolmogorov-Smirnov Test	33
Multi-response Permutation Procedures	40
SUMMARY	54
REFERENCES	56

FIGURES

<u>Number</u>		<u>Page</u>
1	Illustration of topics discussed in the text	2
2	Bar graph of brook trout spawning depths. Depth interval = 0.1 feet	7
3	Bar graph of brook trout spawning depths. Depth interval = 0.2 feet	8
4	Bar graph of brook trout spawning depths. Depth interval = 0.2 feet	9
5	Example demonstrating the relation between a box plot and stem-and-leaf display	15
6	Box plot of spawning depth of 57 brook trout with outliers plotted individually along the depth scale	17
7	Focal point velocity bar graphs representing data collected by three different methods for adult white suckers	18
8	Box plots for distribution over depth of white sucker adults found by three different methods	19
9	Water velocity for spawning trout	20
10	Frequency and empirical cumulative frequency distributions of mean column velocity used in the daytime and nighttime by juvenile white suckers	38
11	Empirical cumulative frequency distributions of daytime and nighttime mean column velocity used by juvenile white suckers	39
12	Sample data for demonstration of multi-response permutation procedures	45
13	Two of the possible and the observed permutations of Table 13 for the observations shown in Figure 12	50
14	Box plots for focal point and mean column velocity for juvenile white suckers as determined by three methods of observation	52

TABLES

<u>Number</u>		<u>Page</u>
1	Habitat values for white sucker adults	5
2	White sucker adult habitat data arranged and summarized by fish focal point velocities	5
3	Calculation of two kinds of relative frequency for the data in Table 2	6
4	Velocity observations on 57 spawning brook trout	10
5	Stem-and-leaf display of velocity values given in Table 4	11
6	Stem-and-leaf display of velocity values from Table 5 with the leaves sorted	12
7	Stem-and-leaf display of data from Table 6 with twice as many lines	13
8	Presence and absence of rock cover for adult white suckers sampled by three different methods	21
9	Percent presence and absence of rock cover for adult white suckers sampled by three different methods	21
10	Simple data to illustrate the Mann-Whitney-Wilcoxon two- sample test	27
11	Distances between all possible pairs of observations in Figure 12	46
12	All pairs of distances within each group for the first five possible permutations into two groups of sizes 3 and 4 of the observations shown in Figure 12	47
13	All possible permutations of the observations A_1 - A_3 and B_1 - B_4 shown in Figure 12	48
14	Comparison of statistical analyses by four methods on several sets of data	51

ACKNOWLEDGMENTS

I thank Bob Moody and David Hansen for providing the data used for illustrations. I also thank the seven reviewers of an earlier version of this manuscript. They provided useful comments and kept me from making several mistakes. I especially thank Dr. Paul Mielke for his encouragement of my presentation of the multi-response permutation procedures. Dr. Mielke also has given permission to the National Ecology Research Center to use and distribute his computer programs for the computation of these procedures.

INTRODUCTION

The Instream Flow Incremental Methodology (IFIM) uses knowledge of species' habitat requirements to quantify habitat available at different stream flows. These analyses may serve many purposes but always depend on specifying a relation between the species and habitat factors of concern. The species-habitat relation expresses differences in the suitability of different values of a habitat factor. Habitat requirements may be given in the form of a suitability index (SI) or a suitability index curve. These curves also are referred to as species criteria (Armour et al. 1984; Bovee 1986) when they function as design criteria in IFIM models used to help recommend stream flows.

Data collection for curve development is time consuming and expensive; therefore, analytic techniques that make full use of the data need investigation. The concern of this paper is to describe a variety of graphic and statistical techniques for exploring and comparing suitability curve data. The habitat factors of depth, velocity, substrate, and cover are emphasized, since they are important in habitat versus stream flow analysis, but the techniques apply to curves for any species or life stage for any habitat factor, aquatic or terrestrial.

An array of such techniques will help fulfill many purposes. In a given study it may be desirable to know if two or more species, or life stages of a species, have similar enough curves to combine in a single instream flow analysis. It is also important to know if, for a particular species or life stage, daytime and nighttime curves or winter and summer curves are the same or different. Data collected in different rivers can be compared to address the question of the generality or transportability of a set of curves. Also, questions about the classification of species into guilds based on similar habitat requirements cannot be addressed without viable techniques to compare curve data.

My discussion of suitability curve data comparison will cover the topics shown in Figure 1. An investigator, with suitability curve data in hand, begins to explore curve data by describing and summarizing the data for each curve of interest. These exploratory steps are mostly graphical and let the investigator see the relations between or among the curves being compared. Such preliminary steps, even though not part of formal hypothesis testing, are valuable for forming hypotheses, detecting errors, and generally for getting comfortable with the data. For some data the graphical methods may be sufficient for concluding that the batches are the same or different.

The techniques I cover in this exploratory, descriptive phase are easy to do by hand and also are available in many statistical program packages. I

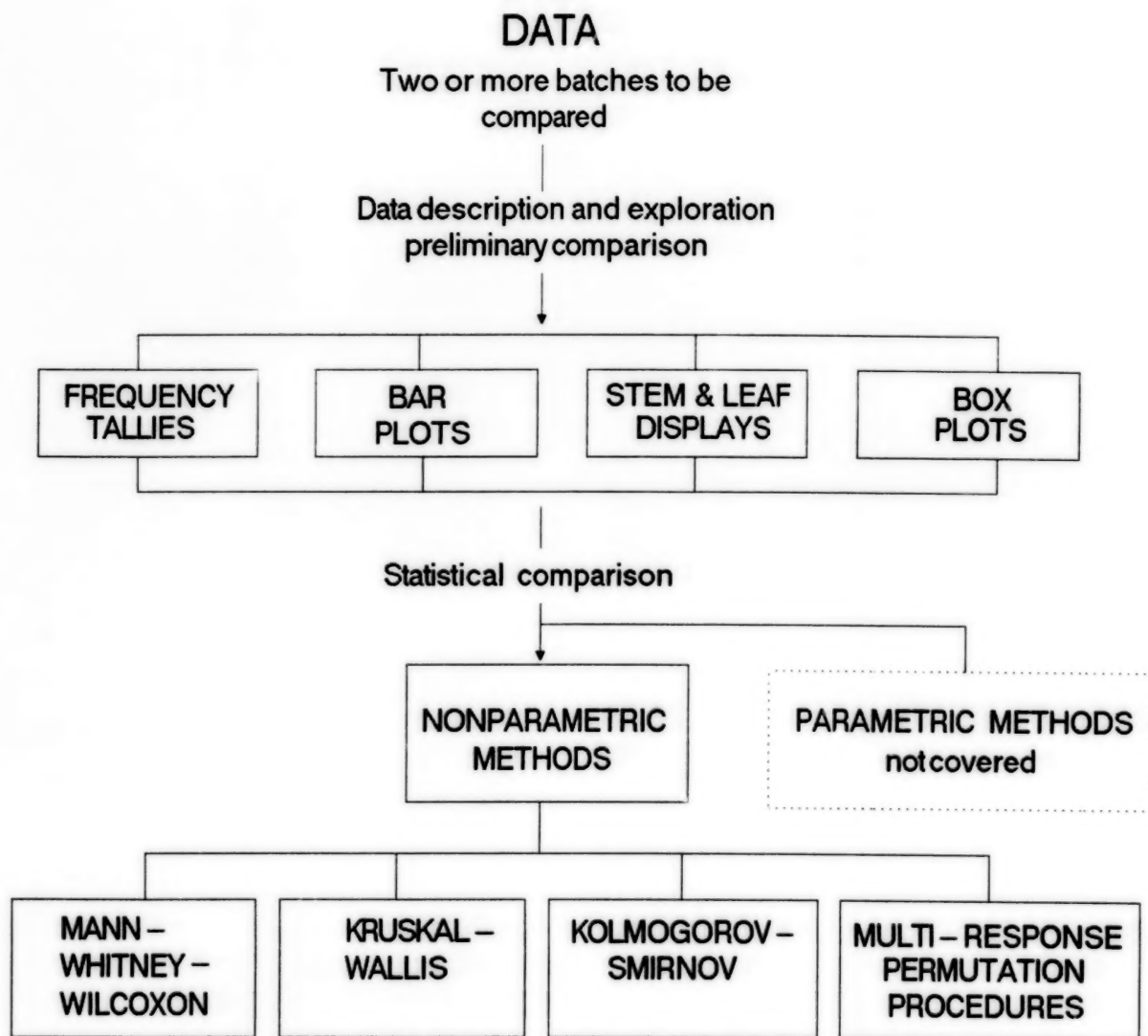


Figure 1. Illustration of topics discussed in the text. The discussion proceeds from top to bottom and left to right. An investigator with two or more batches of habitat suitability curve data should perform some or all of the data description procedures named in the top row, then choose the appropriate test procedure named in the bottom of the figure.

urge that these steps not be omitted; it is all too easy, especially when using a computer for analysis, to immediately perform some statistical procedure thought to be appropriate, but which is not. Indeed these techniques promote familiarity with the data that, before computers, an investigator gained in passing by manually copying and plotting data.

Next I discuss several ways to statistically compare two or more batches, or samples of habitat suitability data (bottom of Figure 1). The rationale, assumptions, and interpretation of each method are described. A complete example demonstrating the method is also given. In general, habitat suitability curves are compared by using the original, unsmoothed data of response to habitat by a species or life stage. The scatter (variability) of the data allows statistical evaluation of the similarities or differences between or among samples. Samples can be compared with respect to how high or low their values are (their location or central tendency), how spread out or concentrated they are (their spread or scale), and how lopsided or symmetrical they are (their skewness or shape). A cautionary note: do not confuse the statistical meaning of "location" as position on a scale with geographic location. Samples from different locations may have the same central tendency or location on a habitat variable.

Since species' curves typically do not approximate a bell shape (the normal curve), I emphasize methods of analysis that do not depend on the normal or any other kind of distribution for the data. Methods that do not depend on a particular distribution are called "distribution free" or "nonparametric." These nonparametric methods are more powerful than their parametric counterparts when data do not meet the assumptions of normality (Conover 1980; Zar 1984). Nonparametric methods often are easier to compute and their rationale easier to understand. Further, many of the methods I discuss are not overly influenced by a few outliers or wild data points, thus they are also called "resistent," "robust," and "sturdy" statistics (Mosteller and Rourke 1973).

EXPLORING AND DESCRIBING CURVE DATA

The usual data for suitability curve comparison is a sample or batch of data for each curve to be compared. Each batch contains the response of a species or life stage to different values of a habitat factor. Species' response is often represented by the number, or frequency, of organisms occurring in a sample of a given range of a habitat variable, but also can be expressed as population density, productivity, or biomass.

Data for curve development are collected in different ways (see Bovee and Cochnauer 1977; Bovee 1986). Commonly the water depth and velocity and channel characteristics, such as substrate and cover, are measured or noted at the locations where fish were observed. Each recorded observation represents a fish and consists of the values of the measurements taken. The habitat features measured (e.g., depth, velocity) are called "variables" or "habitat variables." Sometimes the data are adjusted or collected to reflect habitat preference rather than habitat use. I take the term "habitat suitability" to be ambiguous between use and preference; the methods I discuss apply to either type of data. I presume, in discussing the various methods, that the data are adequate to the purpose of describing species' suitability curves, that they were fairly collected in an appropriate sampling scheme, and that they were correctly recorded, transcribed, and entered into a computer. I discuss some

data checking procedures and describe the sampling requirements assumed by the tests. I am not discussing methods to rescue bad data, but to follow good data through good analyses.

A sample set of data used to illustrate some of the methods described below is shown in Table 1. This is part of a larger data set and is typical of data used in habitat suitability curve development. The data are for the response of white sucker adults to the habitat variables of depth, focal point velocity, dominant substrate particle size, and the presence of rock cover. The data appear much as they would if transcribed from a field book. If analysis is to proceed by computer they can be entered directly. However, for hand analysis it is convenient to order the data by values of each habitat variable and report the number of observations for each different habitat value. Notice that only certain values occur for each variable in Table 1. For example, depth is recorded to the nearest centimeter and velocity to the nearest tenth of a foot per second because this is the accuracy of measurement.

The velocity data from Table 1 are presented in ordered form in Table 2. Next to each distinct velocity value is recorded the number of observations (fish) found at that velocity. The number of observations is also called the frequency and is easy to produce from data arranged as in Table 1 by first listing the velocity values in order, then making a tally, as in Table 2, while reading down the velocity column of Table 1. Also indicated in Table 2 are the implied limits surrounding each velocity value. Here they represent the implied accuracy of measurement but are also a ready reference for plotting a bar graph of the data, since they indicate the positions of the edges of the bars. The measured velocity values will be represented by the midpoints of the bars.

Data arranged as in Table 2 gives a first impression of the shape of the species' suitability curve. Here the fish seem to be concentrated at low velocities, perhaps tailing off at 0.5 feet per second, and not occurring at higher velocities. I recommend that the first step in analyzing suitability curve data is to produce by hand or computer such a frequency plot, bar graph, or other similar form of visual display of the data, even before computing the mean, median, range, etc.

When batch sizes are roughly the same, the resulting frequencies are of the same scale and easily compared visually. But if one sample is substantially larger than another, displaying frequency values or tallies as in Table 2 may not be adequate. In such cases the data can be comparably scaled by calculating and then plotting the relative frequency rather than the raw frequency (Table 3). This is done in one of two ways: each frequency value can be expressed as a percent or proportion of the sample size or as a percent or proportion of the largest raw frequency value. Relative frequency is calculated by dividing each raw frequency value by either the sample size or the largest frequency value. Either method gives a value in the range 0.0 to 1.0. If division is by the sample size then the sum of the relative frequencies equals 1.0; if division is by the maximum frequency then the largest relative frequency value is equal to 1.0. These procedures, especially the second, are sometimes referred to as normalizing or standardizing the data; neither should be confused with standardization by mean and standard

Table 1. Habitat values for white sucker adults. Velocity is focal point velocity. Dominant substrate codes are 4 = sand, 5 = gravel, 6 = rubble. Rock cover is indicated as present = 1 and absent = 0.

Observation number	Depth (cm)	Velocity (f/s)	Dominant substrate	Rock cover
1	62	0.1	5	1
2	61	0.0	5	1
3	103	0.1	4	0
4	83	0.0	4	0
5	67	0.2	5	0
6	80	0.0	4	0
7	96	0.2	4	0
8	80	0.1	4	0
9	72	0.5	6	1
10	26	0.2	4	0
11	47	0.2	4	0
12	47	0.3	4	1
13	50	0.2	4	1

Table 2. White sucker adult habitat data arranged and summarized by fish focal point velocities. Typically the tally is made while reading Table 1, then entering the count in the frequency column.

Implied class limits (f/s)	Velocity (f/s)	Frequency	Tally
0.0 - < 0.05	0.0	3	XXX
0.05 - < 0.15	0.1	3	XXX
0.15 - < 0.25	0.2	5	XXXXX
0.25 - < 0.35	0.3	1	X
0.35 - < 0.45	0.4	0	
0.45 - < 0.55	0.5	1	X

Table 3. Calculation of two kinds of relative frequency for the habitat data in Table 2. The proportions are rounded to two decimal places. Percent relative frequency can be calculated by multiplying the relative frequency by 100.

Velocity (f/s)	Raw frequency	Relative frequencies	
		Raw / sum-of-raw	Raw / maximum-of-raw
0.0	4	4 / 13 = 0.31	4 / 5 = 0.80
0.1	2	2 / 13 = 0.15	2 / 5 = 0.40
0.2	5	5 / 13 = 0.38	5 / 5 = 1.00
0.3	1	1 / 13 = 0.08	1 / 5 = 0.20
0.4	0	0 / 13 = 0.00	0 / 5 = 0.00
0.5	1	1 / 13 = 0.08	1 / 5 = 0.20
	13	1.00	

deviation. Either kind of relative frequency can be expressed as a percent by multiplying by 100. Note that relative frequencies are mainly used to facilitate inspection of the data; statistical tests typically should use the original values.

For large batches or batches scattered over a wide range of a habitat variable, such tallies or bar graphs may be cumbersome to produce or may contain enough noise to obscure underlying pattern. One possibility is to combine two or more adjacent intervals into a single interval or bin and sum the frequencies for the new, wider bin. A bar graph of the depth used by 57 spawning brook trout is shown in Figure 2 with bar widths equal to the measurement scale (0.1 ft). The right tail is somewhat jagged, but the trend of finding fewer fish at greater depths is moderately clear. What is not clear is the shape of the response in shallower water. For example the single highest bar at 1.0 feet is next to an empty bar at 1.1 feet and a bar representing one fish at 0.9 feet. Also, values seem to be heaped up at convenient values like 1.0, 1.5, and 2.0 perhaps indicating rounding bias in the way depths were measured. Combining adjacent bars, shown in Figure 3, reveals a consistent high use of depths below 1.0 feet. The trailing off to the right however is still jagged.

Notice that bins made up of an even number of original intervals have midpoints halfway between two of the original habitat values. Thus the first bin containing the number of fish observed at 0.3 and 0.4 feet is centered at 0.35 feet. This seems a simple point, but it is surprising how confusing plotting bin midpoints and edges can get. I find it easiest to plot the midpoints of the original intervals between the lines of graph paper and draw

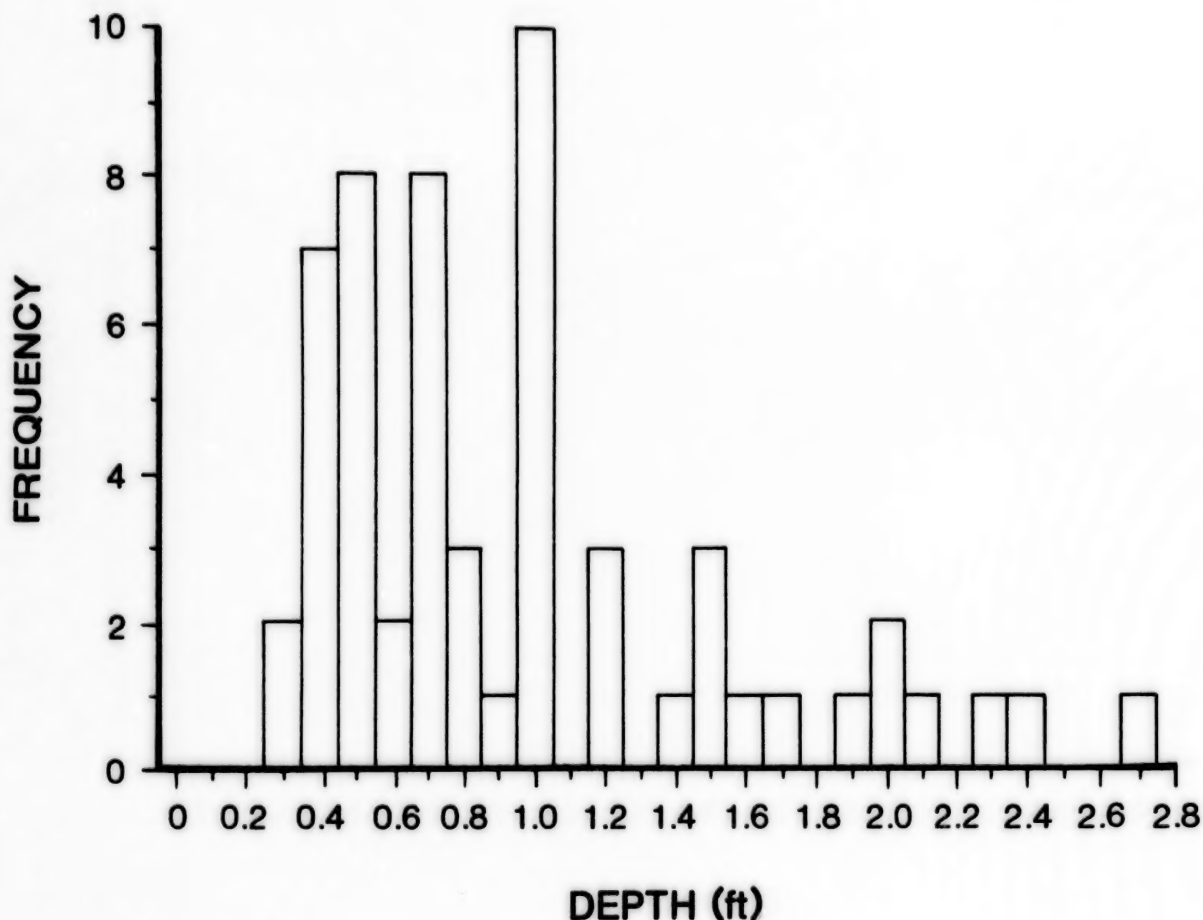


Figure 2. Depths used by 57 spawning brook trout. Bars represent intervals 0.1 feet wide and are centered at the measured depths. Frequency is the number of fish observed at each depth interval.

the bin boundaries along the lines because bin boundaries will always coincide with the outside boundaries of the original intervals for each bin.

Another way in which combining bins can be confusing is that when intervals are combined in pairs, two different pairing arrangements are possible. To get Figure 3, I combined pairs of bars in Figure 2 from left to right. Figure 4 shows the same data but this time pairing started on the right; the bins combined are shifted 0.1 feet. Now the left-most bin has its midpoint at 0.25 feet. Of course if bins 0.3 feet wide are constructed, then there are three possible ways of combining original intervals, four ways to make 0.4 foot wide bins and so on. What is revealing about comparing Figures 3 and 4 is that in the first the most fish occur in water just less than 1.0 feet deep and in the second most fish occur in water just less than 0.5 feet. This difference is purely artificial; remember Figures 3 and 4 are representations of the same data. Because of such discrepancies, combining intervals is not a useful way to compare curve data.

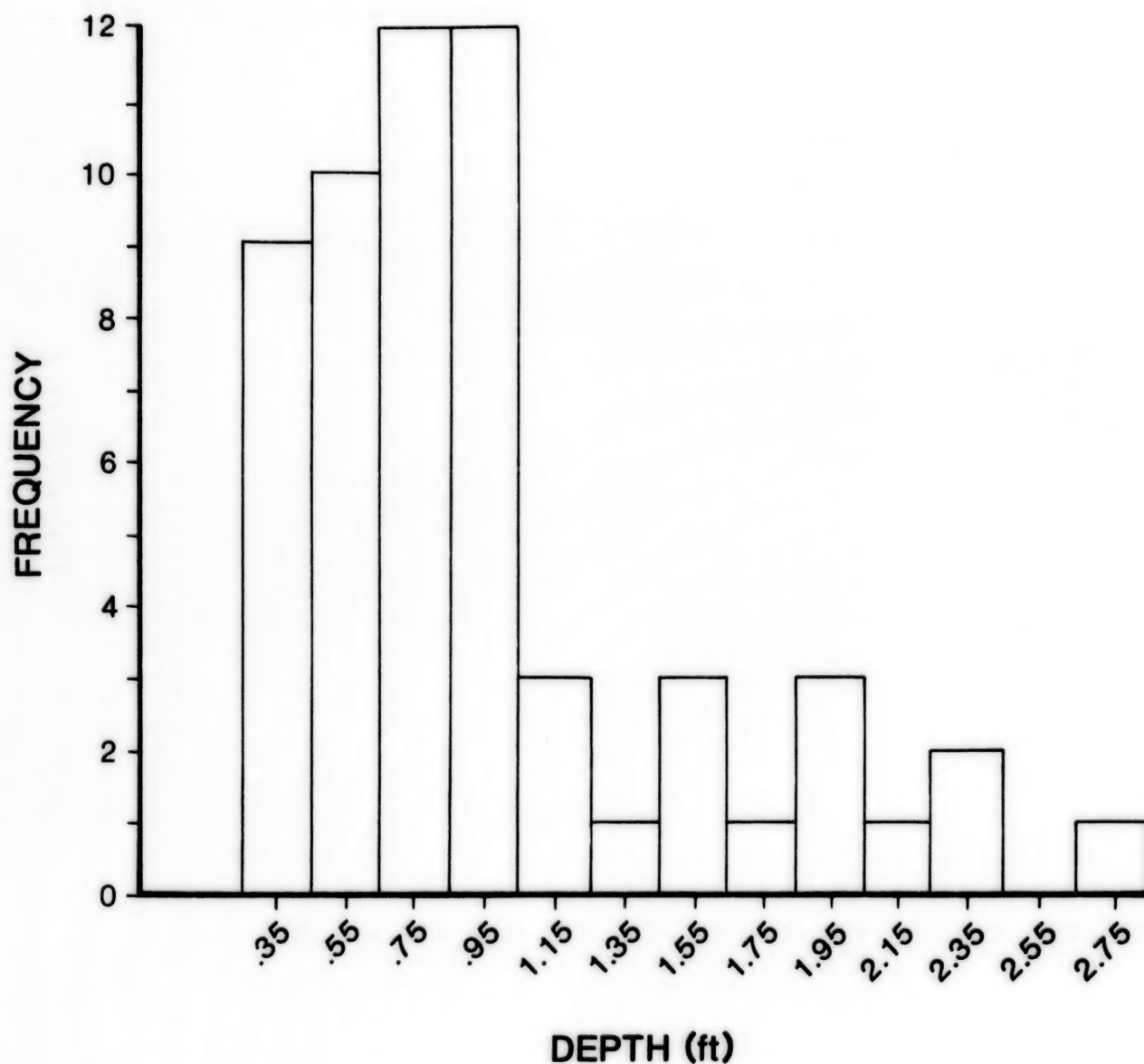


Figure 3. Bar graph, frequency plot, of data from Figure 2. Bars represent intervals 0.2 feet wide with midpoints as indicated. Frequency is the number of fish observed at each depth interval.

Another possibility for summarizing large or scattered batches as an aid to curve comparison is to produce a stem-and-leaf display either by hand or by computer. Stem-and-leaf displays combine the graphic features of a bar graph or frequency tally with the display of raw numbers possible in a table. They are as easy to produce by hand as bar graphs and contain more information.

Table 4 lists the mean column water velocity measured at the locations of 57 spawning brook trout. Each value in the table is to be represented by a

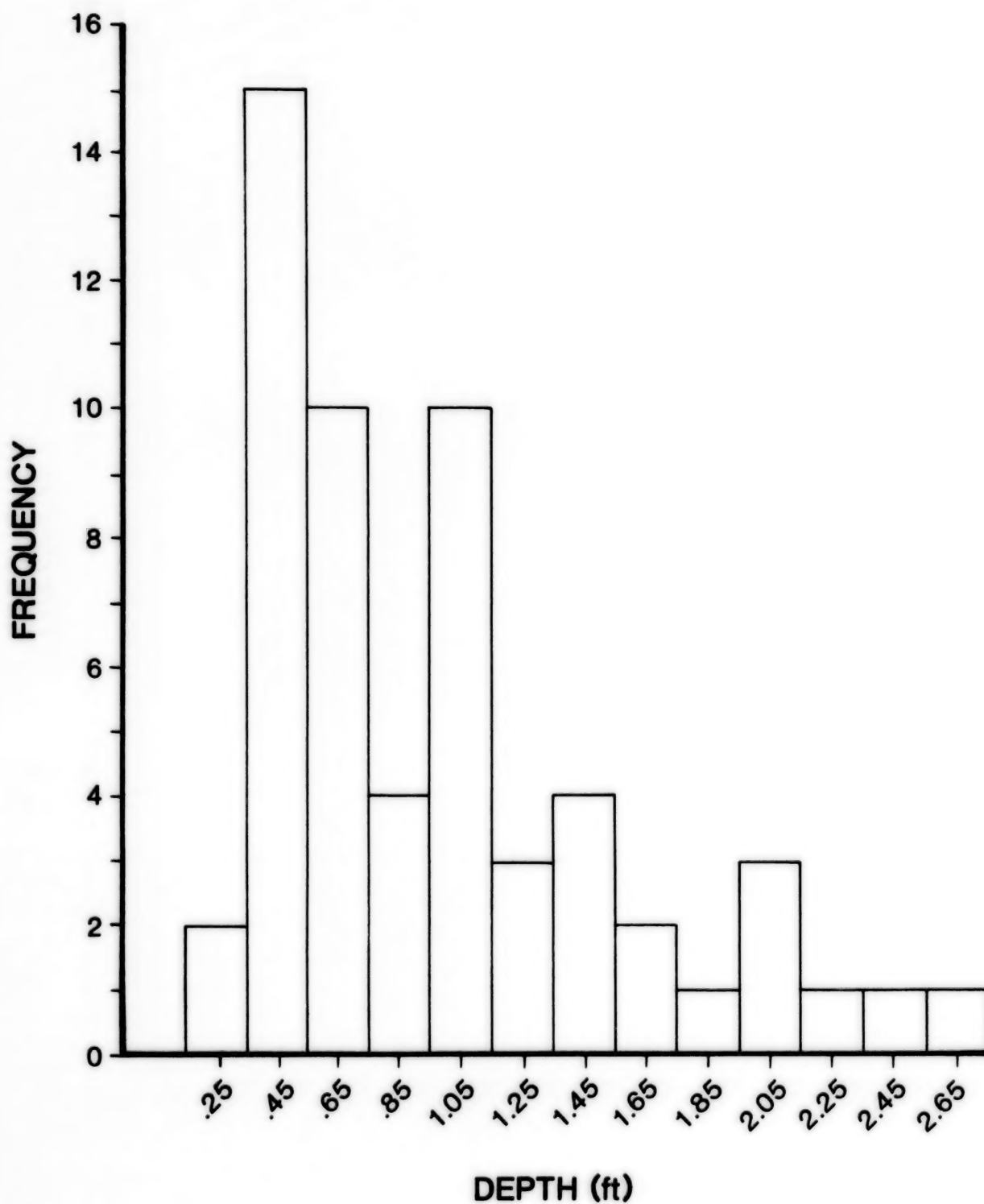


Figure 4. Bar graph, frequency plot, of data from Figure 2. Bars represent intervals 0.2 feet wide with midpoints as indicated. Frequency is the number of fish observed at each depth interval.

Table 4. Velocity observations (feet per second) on 57 spawning brook trout.

0.48	0.00	0.00	0.24	0.10	0.20
0.65	0.00	0.40	0.15	0.05	0.20
0.00	0.20	0.59	0.10	0.07	0.15
0.28	0.37	0.83	0.00	0.02	0.15
0.14	0.68	0.83	0.67	0.15	0.25
0.14	0.00	0.10	0.00	0.15	0.25
0.00	0.00	0.38	0.40	0.00	0.10
0.17	0.45	0.00	0.31	0.10	
0.00	0.17	0.00	0.05	0.15	
0.00	0.00	0.24	0.05	0.15	

stem and a leaf. In this example, use the digits in the tenths place as the stems and the digits in the hundredths place as the leaves. Quick inspection reveals that the values range from 0.00 to 0.83 feet per second; thus the stems are 0.0 to 0.8 and are listed next to the vertical line in Table 5. The leaves are listed to the right of each stem. After listing the stems vertically, the leaves are written in while reading down Table 4. The first value in Table 4 is 0.48 thus the leaf 8 appears as the first leaf next to the stem .4 in Table 5. Continue copying in leaves and soon the shape of the curve emerges, for the stem-and-leaf display is much like a bar graph turned sideways.

Here is more help on reading stem-and-leaf displays. The 15 0's (zeros) in the top line of Table 5 represent the 15 data values in Table 4 that are equal to 0.0. The two 3's next to stem .8 at the bottom of the display represent the two velocity values equal to 0.83 feet per second found near the middle of the third column of Table 4. Thus all the raw data values appear in the display and the batch is nearly sorted. The stem-and-leaf display is also efficient. It takes 171 characters to list the values in Table 4 (228 characters if the leading zeros are needed). The stem-and-leaf display needs only 66 characters, 57 for the leaves and 9 for the stems plus the vertical line. Such efficiency means that raw data may more easily be included in published works and reports.

For many purposes the display in Table 5 is sufficient for inspecting a batch, but revising the table with each line sorted is easy. Table 6 is a fully sorted stem-and-leaf display of the velocity data; also, as is customary, the decimal points in front of the stems are not given and a key showing the relative magnitude of the displayed values is provided.

Several modifications of these displays are possible (Tukey 1977; Velleman and Hoaglin 1981; Chambers et al. 1983; Emerson and Hoaglin 1983). One modification useful to note is to stretch out the display by listing each stem

Table 5. Stem-and-leaf display of velocity values given in Table 4. Construction is explained in the text. A leaf of 5 opposite stem .1 represents a velocity of 0.15 feet per second. There are seven such values in this display because there are seven leaves equal to 5 next to stem .1.

Stems	Leaves
.0	00000000000000555720
.1	4477050055055550
.2	80440055
.3	781
.4	8500
.5	9
.6	587
.7	
.8	33

twice (Table 7). The first stem line takes the leaves 0, 1, 2, 3, and 4; the second takes leaves 5, 6, 7, 8, and 9. This stem-and-leaf display is more choppy than the previous one, but viewing a less smooth display may be valuable. For example, gaps appear in the data for higher velocities indicated by empty lines for stems 5, 6, and 7. Perhaps the two values at stem 8 are outliers, and perhaps the dip in the area with velocities of 0.05 to 0.09 should be investigated. Choosing the number of lines for a stem-and-leaf display is much like choosing bin sizes for bar graphs. Several rules for choosing the number of lines and bin width are given in Emerson and Hoaglin (1983).

The stem-and-leaf displays of velocity also reveal another possible pattern in this batch. Consider the actual values the leaves take; it is curious that the digits 0 to 9 are not represented more or less equally as possible values for the hundredths place as one would expect. Rather, nearly half (25 of 57) are 0, and 14 more are 5. The digit 6 never appears in the hundredths place, and the digits 1, 2, and 9 appear once each. This pattern is significantly different from a random selection of 57 digits and is in need of explanation.

Table 6. Stem-and-leaf display of velocity values from Table 5 with the leaves sorted. Decimal points have been omitted. The smallest value is 0.00 and the largest is 0.83. The key 2 | 4 at the top of the display represents 0.24 feet per second.

Stems	Leaves
2	4
0	000000000000000025557
1	0000044555555577
2	00044558
3	178
4	0058
5	9
6	578
7	
8	33

One explanation perhaps is that velocity had been measured in centimeters per second, converted to feet per second, then rounded to hundredths, but a few calculations shows that this would result in a preponderance of leaves equal to 0, 3, 6, or 9. Another possible explanation might be the way in which the velocity meter was read, especially at velocities near 0.00 feet per second, since many of the leaves equal to zero are on the stem equal to zero. Many readings of 0.00 were perhaps slightly higher, but not recorded that way because of the comparative difficulty of measuring low velocities. The large number of leaves equal to 5 might also indicate some bias, for inspection of Table 4 reveals that 12 of the 14 5's occupying the hundredths place occur in the last half of the list. The observations in Table 4 are listed in the order in which the measurements were made. Perhaps the method of measuring or the person measuring velocity changed during the course of data collection.

My purpose in mentioning these explanations is not to get an answer for this batch, but rather to illustrate how these exploratory, graphic techniques

Table 7. Stem-and-leaf display of data from Table 6 with twice as many lines. The first line of each stem takes leaves 0-4, the second takes leaves 5-9. 2 | 5 represents 0.25 feet per second.

Stems	Leaves
0	0000000000000002
0	5557
1	0000044
1	555555577
2	00044
2	558
3	1
3	78
4	00
4	58
5	
5	9
6	
6	578
7	
7	
8	33
8	

may lead to finding errors, quirks, or bias in the data by revealing patterns not obvious in more standard forms of data presentation. To answer a question posed earlier, the dip in velocity shown at the top of Table 7 might well be artificial. A stem-and-leaf display of a batch is especially useful for inspecting data, since it preserves the raw values as a table yet abstracts the shape of the batch as a bar graph.

Another way to display the shape of a batch, especially when there are many observations or the data show lots of scatter, is to make box-and-whisker plots (Tukey 1977; Velleman and Hoaglin 1981). These box plots divide the data (approximately) into fourths. The first fourth contains the first 25 percent of the observations along the habitat variable, the second the next 25 percent, and so on. Box plots are useful for showing location, spread, skewness, tail lengths, and the presence of outliers.

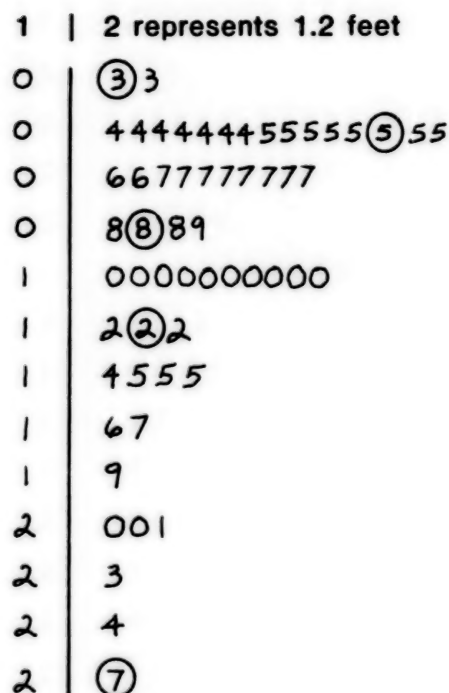
Box plots are easily produced, since the positions of only five of the observations on the habitat variable are involved. These are the two extreme values, the median or middle value, and the values half way between the median and each extreme value. Locating the five points is easy if the data are sorted such as in a stem-and-leaf display. Note that box plots are constructed from raw frequencies; they cannot be constructed from relative frequencies alone. Box plots are available in several of the standard statistical computer program packages. Program listings in BASIC and FORTRAN for constructing box plots (also for stem-and-leaf displays) are given in Velleman and Hoaglin (1981) along with addresses for getting the programs in machine readable form.

Figure 5 shows the steps for making a box plot from a stem-and-leaf display of the depths for 57 spawning brook trout. The five needed values are circled and are determined as follows. The first and last leaf in the display are the extreme values and represent values of 0.3 and 2.7 feet. The median is the third necessary value. It is the observation that divides the batch exactly in half; that is, there are just as many values above it as below it. Since there are 57 values the median is the 29th sorted value and is calculated thusly: $(57 + 1)/2 = 29$. By counting in from either end of the display the 29th value is found to be 0.8 feet. Do not mistake the median, 0.8, with the count of the median, 29. When counting leaves, count right to left if you are counting on lines from the bottom up and count left to right if you start at the top.

For samples with an even number of observations there is no middle value so the median is defined to be the average of the two values closest to the middle. For example, in a batch of 50 the depth of the median in the sorted list of data is $(50 + 1)/2 = 25.5$ or half way between the 25th and 26th values. The median is the average of the values of the 25th and 26th observations.

Next the values half way between the median and the extremes are found. They are commonly called the hinges and can be thought of as the medians of the lower and upper halves of the batch. Thus they are the 15th values in from each end of the sorted batch $((29 + 1)/2 = 15)$. In this example the lower hinge equals 0.5 feet, and the upper hinge equals 1.2 feet. If the

STEM-AND-LEAF DISPLAY



BOX PLOT

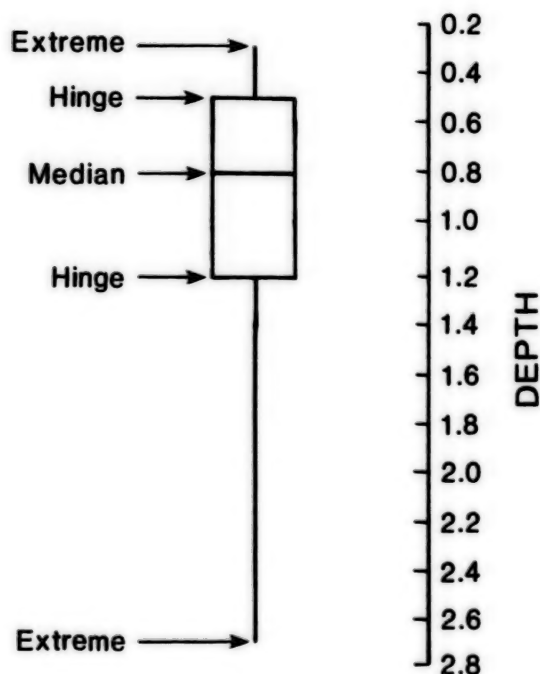


Figure 5. Example demonstrating the relation between a box plot and stem-and-leaf display. The data are the depths for 57 spawning brook trout. The five critical observations are circled in the display to the left and used to construct the box plot to the right. The depth scale implied by the stem-and-leaf display is shown on the right.

median is between two observations (e.g., at a count of 25.5) then compute the count of the hinges from the count of the median ignoring the fractional part (e.g., $(25 + 1)/2 = 13$). Count this number of values in from the extremes of the data to locate the hinges.

The box plot is drawn along the one dimensional measurement scale as shown in Figure 5. The middle 50 percent of the batch is represented by the rectangle whose short sides are positioned at the hinges. The bar across the rectangle is positioned at the median. The first and last quarter of the batch are represented by the lines (whiskers) extending from the rectangle to the extremes. The box plot was drawn vertically to show its relation to the stem-and-leaf display, but it could as easily been drawn horizontally.

The hinges and median are sometimes referred to as quartiles, since they divide the batch into quarters. The distance between the hinges is called the

H-spread (short for half-spread and hinge-spread) or the interquartile range (IQR). The width of the rectangle between the hinges is arbitrary, but can be scaled relative to the sample size (McGill et al. 1978).

The box plot in Figure 5 reveals that the batch ranges from 0.3 to 2.7 feet. The batch is centered around 0.8 feet as indicated by the bar for the median. The H-spread is the difference between the upper and lower hinges or $(1.2 - 0.5) = 0.7$. The values are skewed toward deeper water as indicated by the longer whisker on that side.

Outliers are values much smaller or much larger than the bulk of the data. Some outliers indicate measurement, recording, or copying errors. Such errors should be corrected if possible or else removed from further analyses. Outliers also can be a legitimate part of the data. Even so, their presence should be known because some methods of analysis are strongly influenced by extreme values.

A slight modification of the box plot is used to reveal outliers. How far away from the bulk of the data a value must be to be considered an outlier is arbitrary. A commonly used guide is presented by Velleman and Hoaglin (1981). Any values beyond the hinges more than 1.5 times the H-spread are considered outside values. Any values beyond the hinges more than three times the H-spread are far outside values. The cutoffs for outside and far outside values are called inner and outer fences.

Figure 6 presents the box plot from Figure 5 constructed to show outliers. The single lines (whiskers) are drawn to the most extreme value not beyond the inner fences. Any outliers are plotted individually, usually with different symbols for outside and far outside values. In this example, 1.5 times the H-spread is $(1.5 \times 0.7) = 1.05$, making the cutoff for outside values 1.05 units beyond the upper hinge $(1.05 + 1.2) = 2.25$ and 1.05 units below the lower hinge $(0.5 - 1.05) = -0.55$. Thus the values 2.3, 2.4, and 2.7 are individually plotted in Figure 5. No values in this batch are outliers on the left, and none are far outside to the right.

Another enhancement of box plots gives notched box plots. These are box plots that show the approximate 95 percent confidence interval about the median (McGill et al. 1978; Velleman and Hoaglin 1981). In Figure 6 the notch is bounded by parentheses. Sometimes the box itself is drawn to be narrower, or notched, in the region of the confidence interval, hence the name. The position of the notch is calculated by adding to and subtracting from the median the quantity $1.58 (IQR/n^{0.5})$, where IQR is the interquartile range, n is the sample size, and the superscript, 0.5, means to take the square root of n . The constant, 1.58, derives partly from theoretical and partly from empirical results (McGill et al. 1978; Velleman and Hoaglin 1981). A pair of box plots with notches not overlapping can be said to be significantly different at roughly the 5 percent level.

Box plots are good for showing the location, spread, skewness, and outliers for a batch of data. Further, box plots are stable or resistant to quirks in the data. The hinges and median will not change if nearly all the

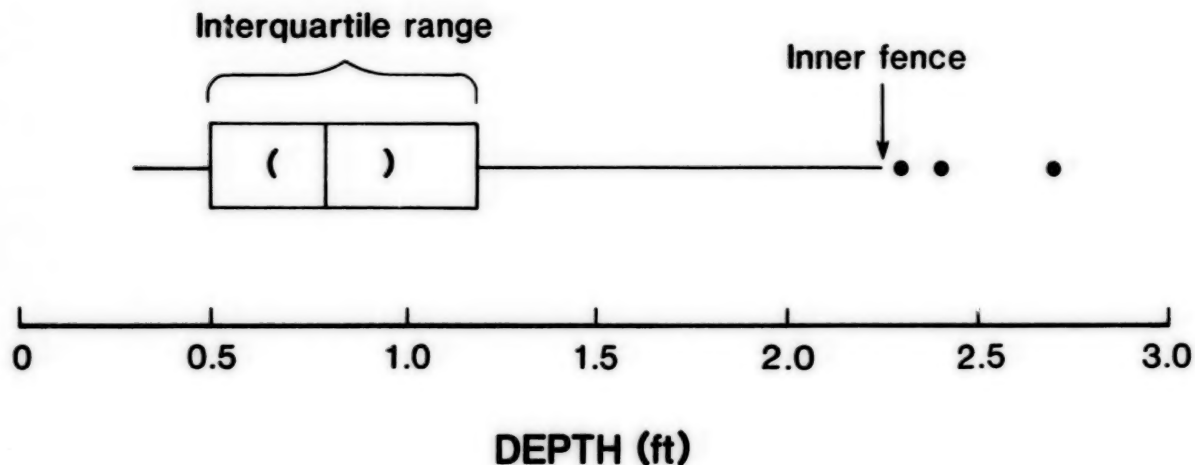


Figure 6. Box plot of spawning depth of 57 brook trout with outliers plotted individually along the depth scale. The approximate 95% confidence interval about the median is indicated by parentheses.

values in the lower fourth are made arbitrarily small and the values in the upper fourth made arbitrarily large. Even the fences will remain relatively stable, since they are defined in terms of the hinges and medians (Emerson and Strenio 1983). Similar plots could be constructed using the mean and some multiple of the standard deviation rather than the median and H-spread, but the mean and standard deviation can change drastically with the occurrence of even a single outlier. Such plots would lack the stability of box plots.

After this preliminary manipulation, graphing, and inspection of the data, curve development can proceed by whatever technique the investigator chooses. But to compare the data from which curves are derived, the next step is to inspect a side-by-side display of frequency tallies, bar graphs, stem-and-leaf displays, or box plots of the samples to be compared. Data for adult white suckers collected by three different methods are displayed in Figure 7. The first sample was collected by electrofishing (these are the data from Tables 1 and 2), the second by visual observation while walking along the stream, and the third by underwater observation. The electrofishing and visual observation data appear similar; both are humped-up at low velocities, with a single observation at a higher velocity. The underwater data, however, appear flat compared to the other two. Whether these appearances indicate substantial similarities or differences will be addressed later when formal or statistical data comparison techniques are presented.

A better way to visually compare curves, especially when there are many observations or the data show lots of scatter, is to draw box plots side by side on the same scale. Box plots for data collected by three methods for white sucker adults for the habitat variable of depth are shown in Figure 8. As with velocity (Figure 7), electrofishing and visual observation appear to

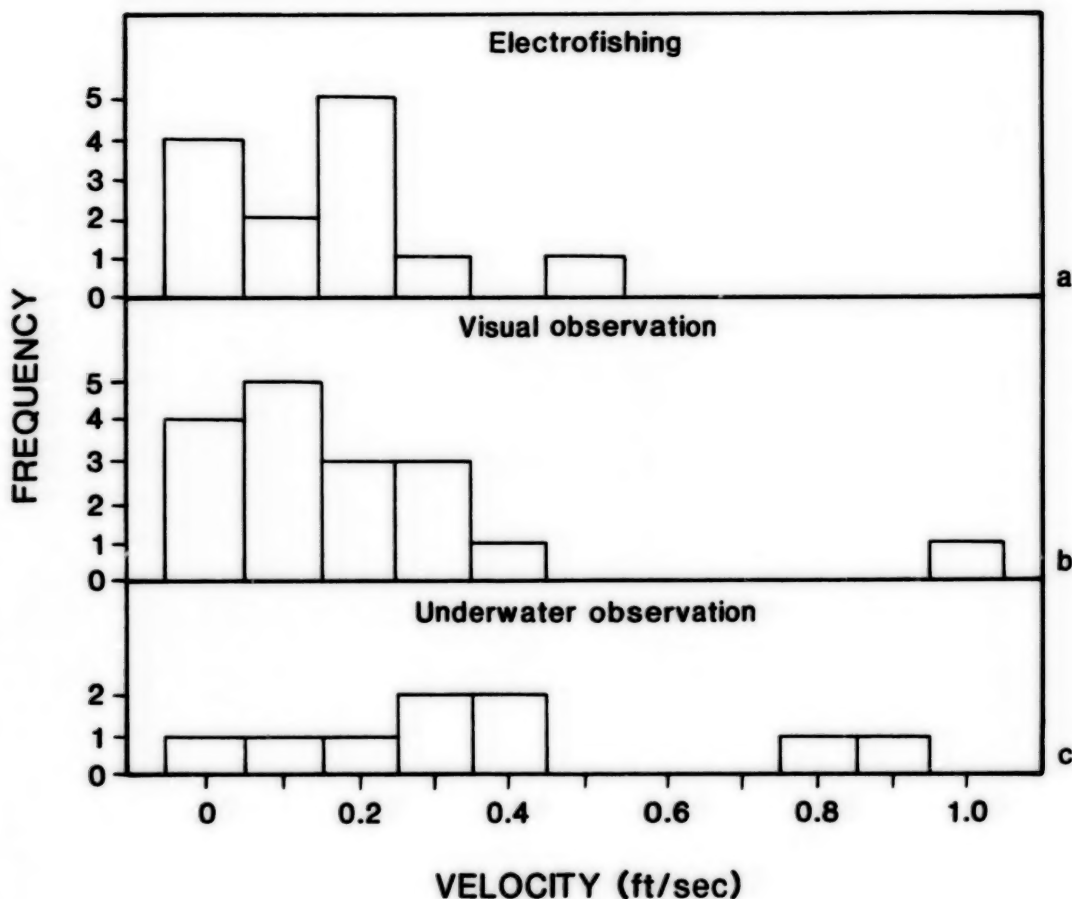


Figure 7. Focal point velocity bar graphs representing data collected by three different methods for adult white suckers.

give similar results, but results from underwater observation are apparently different. That underwater observations show fish to be using deeper water is not surprising here because this technique was rarely used in water less than 1 m deep, and the other methods rarely used in water greater than 1 m deep. The difference in the box plots seems so great and the reason for the differences so obvious that an investigator might well conclude that the methods give different results without performing any statistical tests. Such can be the value of these preliminary steps in data description. Indeed, box plots should be viewed early in the analysis of any grouped data.

Box plots also allow quick visual scanning of several batches at once. Water velocity data for spawning trout are grouped in several ways (Figure 9). The top four plots are for brook trout, the bottom four for brown trout. Half the plots for each species represent individuals that live in a lake and spawn in a stream; the other half represent stream-dwelling fish. Also, for each

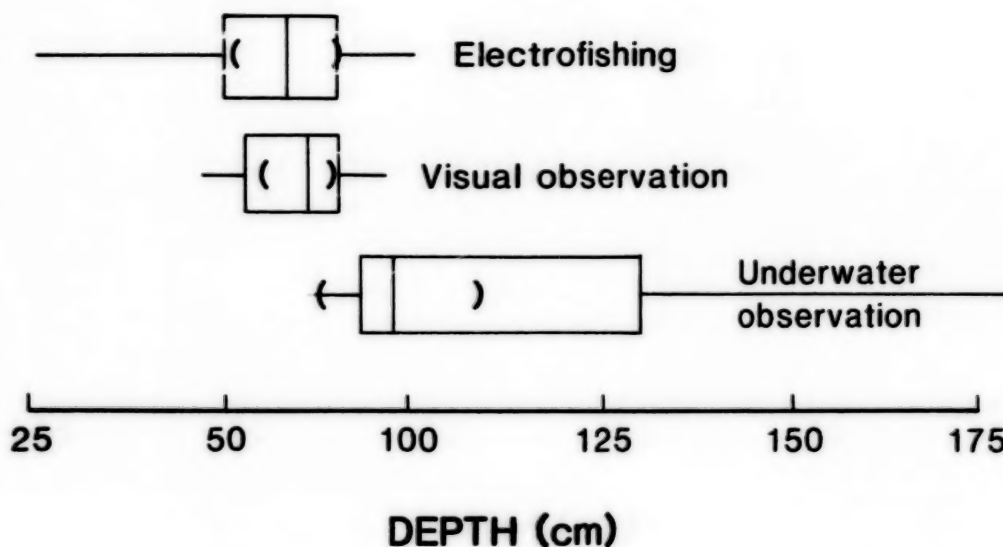


Figure 8. Box plots for distribution over depth of white sucker adults found by three different methods.

combination of species and dwelling location, the lower plot represents mean column velocity, and the upper represents focal point velocity. Several comparisons are readily observable for this relatively complex three-way design. Brown trout generally spawn in faster water, with lake-dwelling brown trout using the fastest velocities, and lake-dwelling brook trout perhaps the slowest velocities. Focal point velocity tends to be lower than mean column velocity, but perhaps not significantly so in some cases. Also, several outliers are noticeable, and the distributions have different ranges and H-spreads.

So far I have discussed habitat variables such as depth, velocity, and substrate (measured in particle size) that have ordered values. However, some habitat variables have values with no particular order, that is, the values represent categories. Substrate categories may include, for example, detritus, sand, cobble, and bedrock. Cover categories may include rock, log, bank, and overhanging vegetation. Cover may also be dichotomous, that is, having only two values. An example is the presence or absence of rock cover for each fish (row) in Table 1. Categorical variables are also called "nominal" scale variables because the categories are nothing more than names. This is so even if the codes used are integers; they act as names, for even though integers are ordered, their assignment to categories is arbitrary.

A different set of methods is used for analysis of categorical data, but its preliminary inspection is much the same. Bar graphs with a bar for each category can be constructed and several such graphs visually compared. It is customary to leave spaces between adjacent bars to indicate that no

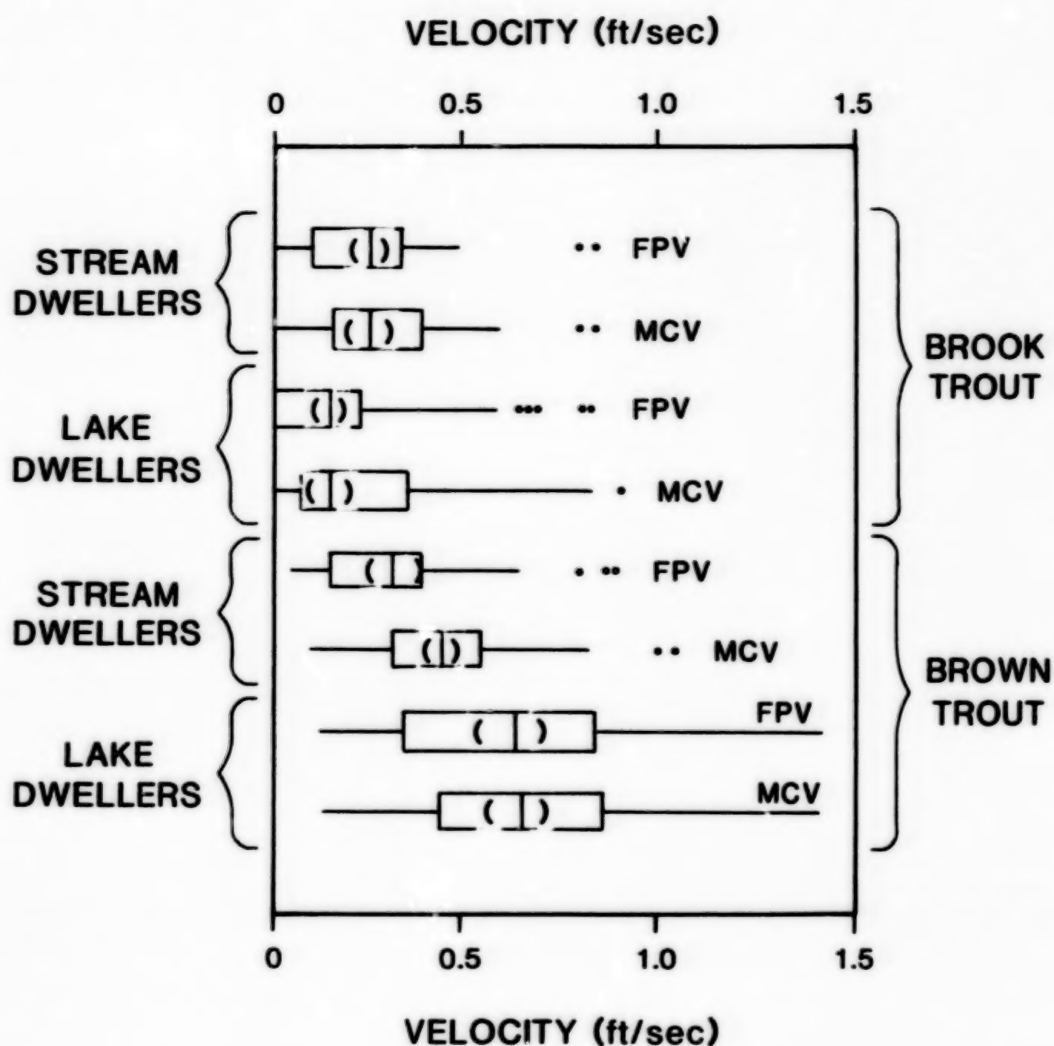


Figure 9. Water velocity for spawning trout. Mean column velocity (MCV) and focal point velocity (FPV) are given for lake- and stream-dwelling brook and brown trout.

intermediate value is possible. Of course box plots cannot be made for categorical data because the categories take no order. To compare two or more batches, categorical data can be listed in a table with a row (column) for each batch and a column (row) for each category. Such an array (Table 8) shows the presence and absence of rock cover for fish found using each of three sampling techniques. Electrofishing and visual observation from the stream bank found fish using and not using rock cover in roughly equal proportions. Underwater observation, however, appears to find proportionately more fish using rock cover. Whether this difference is significant, especially with such small samples, will be investigated later.

Table 8. Presence and absence of rock cover for adult white suckers sampled by three different methods.

Method	Rock cover		Total
	Present	Absent	
Electrofishing	5	8	13
Visual observation	9	8	17
Underwater observation	8	2	10
Total	22	18	40

When sample sizes are different, the presence and absence values can be reported as a percent of the observations for each sample (Table 9). Another aid to inspection of categorical data is to express the data as percents within each category (column percents) or as a percent of the whole table. Most computer programs that analyze contingency tables provide such options.

Table 9. Percent presence and absence of rock cover for adult white suckers sampled by three different methods. Data from Table 8 but expressed as row percents.

Method	Rock cover		Total
	Present	Absent	
Electrofishing	38	62	100
Visual observation	53	47	100
Underwater observation	80	20	100
Total	55	45	100

STATISTICAL ANALYSIS OF CURVE DATA

The next step beyond data description for comparing habitat suitability data is to make statistical inferences about the similarities and differences between or among sets of data. To do this, explicit hypotheses must be stated, the appropriate statistical procedure performed, and the results of the procedure evaluated. Several statistical procedures are described below, but first some general comments are in order.

Statistical tests are roughly classified as either parametric or nonparametric. The most familiar parametric tests are based on the normal curve or the normal distribution and include the t-test and analysis of variance. Parametric tests are so named because they depend on knowledge of, or the ability to estimate, parameters of the populations from which samples are taken. Parameters are properties of populations such as the true mean, variance, and standard deviation. Parameters are estimated from a sample. For example, the population mean is estimated by the sample mean; the sample mean is the sum of all the numbers in the sample divided by the number of values. To legitimately interpret tests that depend on estimation of parameters (e.g., t-test and analysis of variance), the population from which the sample is taken is assumed to have a particular distribution (e.g., normal). That is, an assumption beyond the data in hand is required.

Nonparametric methods, also called distribution free methods, are those that do not depend on knowing or estimating population parameters. Nonparametric methods are distribution free in that they do not depend on any assumption about the distribution of the population in order to make inferences about the population from samples; they depend only on the data in hand. Typically, nonparametric methods are easier to compute than their parametric counterparts, and their rationale is easier to understand.

I will not describe parametric procedures here. Suitability curve data often do not meet the assumptions of parametric methods (e.g., normality), in which case nonparametric methods are usually more powerful. If the assumptions for the parametric tests are met, the parametric tests outperform their nonparametric counterparts, often only slightly. Nonparametric tests are easy to understand and typically easy to compute. In addition, parametric tests are already familiar to many investigators, whereas nonparametric methods typically are inadequately covered in basic statistics courses. Descriptions of parametric tests are abundant; see for example the biological statistics texts by Sokal and Rohlf (1981) and Zar (1984) for discussions of parametric methods and easy-to-follow demonstrations of their computation. Armour and Platts (1983) present parametric methods in the context of monitoring salmonid streams. I will indicate the parametric counterpart of each nonparametric method discussed, if such a counterpart exists.

I said above that to make statistical inferences about the similarities or differences between or among batches of curve data, three steps are taken:

a hypothesis is stated, an appropriate procedure is performed, and the results of the procedure are evaluated. These are discussed in turn.

A statistical inference draws a conclusion about a population from a sample of the population. The conclusion is a statement about the population and is expressed as a hypothesis (e.g., the daytime and nighttime depth distributions of juvenile white suckers are different). What can be confusing is that the hypothesis actually tested is often different from the hypothesis the investigator is interested in. The hypothesis tested is called the null hypothesis, symbolized H_0 , and represents the conditions assumed by the particular test procedure being performed. It stands in contrast to another hypothesis, the alternative hypothesis, symbolized H_1 , which is usually the hypothesis of interest. Most statistical tests are defined to test a null hypothesis expressing conditions of no difference or of the equality of the batches being compared (e.g., the daytime and nighttime depth distributions of juvenile white suckers are the same). A null hypothesis that specifies equality or no difference between two batches is more accurately stated that there are no differences that could not have come about by chance alone.

A reason for this apparently backward, definitely awkward, way to set up statistical tests has to do with the logic of proof in science. A hypothesis is a general statement, and in general such statements cannot be directly proved, only disproved. The strategy of statistical inference is to test a null hypothesis related to the alternative hypothesis of interest such that if the null hypothesis is disproved (rejected), then an alternative must be true. If the null hypothesis is not rejected, however, one cannot conclude that it is true and the alternative false. This is because failure to reject the null hypothesis could be the result of insufficient evidence, poor sampling design, or measurement error. If the null hypothesis is not rejected, then the correct report is that no evidence against the null hypothesis was found.

A null hypothesis expressing no difference between two samples, say in the day and night depth distribution of a fish, is falsified if the daytime depths are either greater than or less than the nighttime depths. A test of such a hypothesis is called a two-tailed test. Sometimes the hypothesis an investigator is interested in specifies the direction of the difference, in which case the null hypothesis is that the samples have no difference or differ in only one direction. (e.g., H_1 : The daytime depth distribution of juvenile white suckers is greater than the nighttime distribution; H_0 : The daytime and nighttime depths are the same, or the daytime depths are less than the nighttime depths.) Here the significance test is a one-tailed test.

After forming an appropriate hypothesis a statistical procedure is carried out to test the hypothesis. A test statistic is calculated from the data in question and used to evaluate the significance of the test. The test statistic for each procedure is calculated in a different way based on the assumptions of the test. For many nonparametric methods the test statistic is not computed from the data values themselves. Rather, values are replaced by their ranks and the test computed using the ranks. The test statistic also may have to be adjusted if, as commonly happens, several data points are tied in rank.

After forming a hypothesis and carrying out an appropriate statistical procedure, the results of the procedure need to be evaluated. For nonparametric methods, the test statistic is often evaluated differently for small and large sample sizes. Usually the test is defined exactly for small sample sizes and approximately for large sample sizes. Statistically rejecting or failing to reject a hypothesis is seldom done with certainty; rather a hypothesis is rejected or not with degrees of probability or confidence. The test statistic derived for a batch is evaluated for the degree of certainty it affords towards rejecting the null hypothesis. The test statistic (for any procedure) computed for a batch is associated with a probability value (p-value). Indeed, this is what makes the procedure a statistical test. The p-value is the probability of having a result as or more extreme than the observed results, assuming the null hypothesis is true.

The statistical reason behind the assignment of p-values is that all the possible values of the test statistic for a particular test can be proven to follow a known distribution, such as the normal curve. Do not be confused; the data need not be distributed normally, rather the test statistic calculated from the data can be evaluated using the normal (or some other) probability distribution function.

Usually the test statistic is compared to a number in a table, but sometimes the significance level can be computed directly. The number in the table is found according to the sample size and the probability or significance level chosen by the investigator. Usually, if the computed value is greater than the value found in the table, then the null hypothesis is rejected and the alternative accepted. The probability or significance level for rejecting a hypothesis should be chosen by the investigator before the analysis is carried out. The probability or significance level is symbolized by the small Greek letter alpha (α). An alpha level (α -level) of 0.05 means that the probability of mistakenly rejecting a true null hypothesis is five percent. Alternatively, and what is becoming the recommended practice, the actual probability value (p-value) produced by a test should be reported, not just whether the probability chosen by the investigator was surpassed or not. This way readers can draw their own conclusions (Warren 1986).

In summary, to make statistical inferences for comparing batches of habitat suitability curve data, proceed as follows.

1. A hypothesis of interest is formulated and taken as the alternative to a null hypothesis, which is the hypothesis actually tested. These hypotheses are so related that if the null hypothesis is indicated as being false, then the alternative is accepted as true. The null hypothesis (usually) states that the batches being compared are not different, that they are samples from the same parent population.
2. A test statistic for an appropriate statistical method is computed from the data. For each statistical method the test statistic is defined such that each of its possible values is associated with a probability (p-value) of getting the observed value or a more extreme value if the null hypothesis is true.

3. The test statistic is evaluated either by directly calculating its p-value or by comparing the observed test statistic to values in a table corresponding to selected probabilities. The p-value is also called the significance level of the test. A low p-value or rejection of the null hypothesis at a low probability level (e.g. $\alpha < 0.05$) indicates that the null hypothesis is (probably) not true.

Nonparametric methods fall roughly into two classes, those that require the data to be ordered and those that do not. Thus one set of methods will apply to data for depth, velocity, and substrate particle size, and another set will apply to the categorical variables substrate and cover. Methods that apply to categorical data often can be used on continuous or ranked data as well because the data can be divided into contiguous classes. However, the methods designed for ordered data will generally be more powerful, since they take into account the ordering. With one exception, categorical methods are not covered here.

Methods suitable for comparing depth, velocity, and substrate particle size data are presented below and include the following: the Mann-Whitney or Wilcoxon test for differences in the location (mean, median) of two independent samples, the Kruskal-Wallis one-way analysis of variance of ranks for differences in the location of two or more independent samples, the Kolmogorov-Smirnov two-sample test for any difference between two sample distributions (location, scale, or shape), and the multi-response permutation procedure (MRPP) for detecting primarily location and shape differences in the distribution of two or more independent samples. In addition, MRPP can be used on dichotomous categorical variables.

These methods assume that the observations within each batch are random and the observations within and between each batch are independent. The requirement of randomness is met if each fish had an equal chance of being observed. Randomness is violated if, for example, fish in deeper or faster water are less likely to be observed than fish in shallower or slower water. This could be so if observations are made from the stream bank. The requirement of independence is met if the value of one observation has no influence on any other observation. Independence within batches is violated if, for example, observing a fish in a deep pool makes it more likely that the next observation will also come from a deep pool. This could be so for fish that tend to aggregate in pools or for observers who spend more time in pools because fish are easier to observe there. Some tests also have other assumptions that will be mentioned as necessary.

THE MANN-WHITNEY-WILCOXON TEST

Two batches of suitability curve data can be compared by the Mann-Whitney test to see if one batch tends to have larger or smaller values than the other. This test is also known as the Wilcoxon two-sample test, which should not be confused with an unrelated technique named the Wilcoxon signed rank test. The null hypothesis is that the two samples come from the same population, thus the samples are identical except for chance differences. The test statistic is computed from the ranks of the observations rather than the

observed values. The ranks are determined by pooling the batches and ranking the combined sample. If the groups come from the same population, then a ranking that puts all of one group ahead of the other group is highly unlikely. Thus if the sum of the ranks assigned to one group is excessively large or small relative to what is expected under the null hypothesis, the null hypothesis is rejected.

The principle behind the Mann-Whitney-Wilcoxon test is that if the samples are from the same population (i.e., the null hypothesis is true), then any combination of ranks assigned to the two samples is equally probable. Consider the simplified data for two groups with sizes 2 and 3 shown in Table 10. For each group the rank order within the combined sample of 5 observations is given. The sums of ranks for the groups are 4 and 11. Whether these are excessively small or large can be addressed by considering all possible assignments of the 5 ranks to two groups of sizes 2 and 3. To do this one need only consider the possible ranks assigned to one group, for if the ranks are specified for one group, the ranks for the other are determined. To further simplify calculation, only the ranks assigned to the smaller group are needed. On the right side of Table 10 are the 10 possible assignments of 2 ranks out of 5 for the smaller group. The possible rank sums thus range from 3 to 9.

If the null hypothesis is true (the two groups represent the same population), then any one of the assignments is possible, and they each have an equal chance of occurring. If the alternative hypothesis is considered (group A is generally smaller than group B), then a rank sum of 4 or less would be possible only 2 times in 10. Therefore a rank sum of 4 or less for these sample sizes will occur by chance 20 percent of the time. If the null hypothesis is rejected when the group A rank sum is 4 or less, then the decision to accept the one-tailed null hypothesis when it is true will be correct about 80 percent of the time (here $\alpha = 0.20$).

A two-tailed test, that group A is smaller or larger than group B, can also be illustrated with the artificial data. Here the decision to reject the two-tailed null hypothesis of no difference can be rejected (at $\alpha = 0.20$ or the 20 percent significance level) if for the smallest group a rank sum of 3 or 9 is found. If this were the hypothesis, then the null would not be rejected for the data in Table 10 since the actual rank sum (4) is greater than 3 and less than 9.

Calculation of the Mann-Whitney-Wilcoxon two-sample test is illustrated in Example 1 for the day and night depth distribution of juvenile white suckers. The null hypothesis is that there are no day/night differences in the depths used. The first step is to pool the data and assign ranks. Tied data values are assigned an average rank as illustrated in the example. The test statistic, U , is computed from the rank sum and corrected for ties if necessary. Tables for samples sizes 20 or less give the exact probabilities of the test statistic (Rohlf and Sokal 1981; Zar 1984). For large samples, a statistic is computed that approximates a normal distribution and can be evaluated by consulting a standard normal table (Sokal and Rohlf 1981). A better approximation of the p-value for this test can be obtained by a variation of the multi-response permutation procedure, which I discuss later.

Table 10. Simple data to illustrate the Mann-Whitney-Wilcoxon two-sample test. The rank order of observations is used instead of original values. Group A has two observations that rank 1 and 3 among the total sample of 5. Also shown are the 10 possible assignments of ranks to group A.

Observed ranks		Possible ranks and rank sums		
Group		Group A		
A	B	Ranks		Rank sums
1	2	1	2	3
3	4	1	3	4
	5	1	4	5
4	11	1	5	6
		2	3	5
		2	4	6
		2	5	7
		3	4	7
		3	5	8
		4	5	9

That approximation uses more of the information in the data. Example 1 follows the presentations by Sokal and Rohlf (1981) and Conover (1980).

THE KRUSKAL-WALLIS TEST

Two or more samples of suitability curve data can be compared by the Kruskal-Wallis test. If only two samples are compared, the Kruskal-Wallis test reduces to the Mann-Whitney-Wilcoxon test. These tests are thus related as are the two-sample t-test and the F-test of analysis of variance. The Kruskal-Wallis test is often called a nonparametric analysis of variance or analysis of variance by ranks.

Example 1. Mann-Whitney-Wilcoxon two-sample test. Data are two samples of the depths used by juvenile white suckers. One set is for daytime use (sample size = $n = 33$), the other is for nighttime use ($m = 17$). The data are arranged below by the value of the depth variable. The average rank and t/T columns are explained below. The sample times are day (D) and night (N).

Rank	Depth (cm)	Average rank (R)	t/T	Time	Rank	Depth (cm)	Average rank (R)	t/T	Time
1	23	1		N	26	53	26		D
2	27	2		N	27	54	27.5	2/6	D
3	28	3		N	28	54	27.5		N
4	30	4.5	2/6	N	29	55	29		N
5	30	4.5		N	30	57	30		D
6	31	6		N	31	58	31.5	2/6	D
7	32	7		N	32	58	31.5		D
8	33	8.5	2/6	N	33	59	33.5	2/6	D
9	33	8.5		N	34	59	33.5		D
10	38	10		N	35	61	35		D
11	41	12	3/24	D	36	64	37	3/24	D
12	41	12		D	37	64	37		D
13	41	12		D	38	64	37		N
14	42	14		N	39	65	39		D
15	43	15		D	40	66	40.5	2/6	D
16	44	16		N	41	66	40.5		D
17	47	18	3/24	D	42	68	42.5	2/6	D
18	47	18		D	43	68	42.5		D
19	47	18		N	44	74	44		D
20	48	20		D	45	75	45		D
21	49	21		D	46	76	46		D
22	50	22		D	47	78	47		D
23	52	24	3/24	D	48	87	48		D
24	52	24		D	49	95	49		D
25	52	24		N	50	109	50		D

The hypotheses are

H_0 : Daytime and nighttime depths are the same.

H_1 : Daytime and nighttime depths are different.

The assumptions are

- 1) Both samples are random selections from their populations.
- 2) The measurement scale is at least ordinal.
- 3) If there is a difference between populations, then the difference is a difference in their locations.

(Continued)

Example 1. (Continued)

Assign ranks.

Combine the two samples into a single list, as above, ordered by the habitat variable. Assign average ranks (R) 1 to 50 ($m + n$). The average ranks for tied depth values (shown in brackets above) are calculated by averaging the ranks that would have been assigned were there no ties. For example the two values of 30 cm have an average rank of $(4 + 5)/2 = 4.5$; the three values tied at 41 cm have an average rank of $(11 + 12 + 13)/3 = 12$.

Compute test statistic.

Arrange the data by group, as below, replacing each original depth value by its average rank (R).

Daytime Average ranks (R)			Nighttime Average ranks (R)	
12	26	40.5	1	10
12	27.5	40.5	2	14
12	30	42.5	3	16
15	31.5	42.5	4.5	18
18	31.5	44	4.5	24
18	33.5	45	6	27.5
20	33.5	46	7	29
21	35	47	8.5	37
22	37	48	8.5	
24	37	49		
24	39	50		

Sample size = $n = 33$

Sample size = $m = 17$

Sum of Ranks = $\Sigma R = 1054.5$

Sum of Ranks = $\Sigma R = 220.5$

Compute the sum of the ranks (ΣR) for the smaller group. The capital Greek letter sigma (Σ) means to take the sum of all the R's in the group. Thus ΣR is the sum of 17 numbers in this case.

$$\Sigma R = (1 + 2 + 3 + 4.5 + \dots + 27.5 + 29 + 37) = 220.5$$

The test statistic is determined in two steps. First use ΣR and the sample sizes (m, n) to compute the value C.

$$C = (m)(n) + \frac{m(m+1)}{2} - \Sigma R$$

$$C = (33)(17) + \frac{17(17+1)}{2} - 220.5$$

(Continued)

Example 1. (Continued)

$$C = 561 + 153 - 220.5 = 493.5$$

Then compare C with $(m)(n) - C = (33)(17) - 493.5 = 67.5$ and use the larger for the Mann-Whitney-Wilcoxon test statistic U . Here

$$(m)(n) - C = 67.5 < C = 493.5 = U.$$

Correction for ties.

If there are ties across samples (ties can exist within one or both samples) and if the largest sample is greater than 20, then compute a correction factor for ties as follows. The values used in the correction factor appear in the column labeled t/T in the combined data table above. For each set of ties between or within groups, note the number of ties (t) in the set. For example, $t = 2$ for the set tied with average rank equal to 4.5, and $t = 3$ for the set of three values tied for rank 12. Now for each set of ties compute a correction factor $T = (t^3 - t)$, which for easier computation can be rewritten as $T = (t - 1)t(t + 1)$. For the set of ties of average rank 12 above $T = (3^3 - 3) = 27 - 3 = 24$ or $T = (3 - 1)3(3 + 1) = (2)(3)(4) = 24$.

The correction factor for ties is the sum of all the T values (ΣT). In this example

$$\Sigma T = (6 + 6 + 24 + 24 + 24 + 6 + 6 + 6 + 24 + 6 + 6) = 138.$$

Test significance of U (small samples).

If there are no (or few) ties across samples (ties can exist within one or both samples) and if the larger sample size is less than 20, then compare U to a value from tables in Zar (1984) or Rohlf and Sokal (1981). The value in the table is found according to the sample sizes and the α -level chosen for the test. If the computed U is equal to or larger than the value in the table, then the test is significant. That is the null hypothesis is rejected at the α probability level. Other published tables may be used, but be warned that some tables for this test are so arranged that the test statistic is significant if its value is smaller than the tabled value.

If there are ties between samples and the largest sample is equal to or less than 20, then comparing the computed U to the tabled value does not yield the p -value in question. The table can be used as described, but will usually result in a conservative inference. That is, the actual p -value will be lower than the α -level chosen from the table.

The sample sizes in this example (33, 17) are too large to use the small sample method, and thus the large sample approximation is appropriate.

Test significance of U (large samples).

If the largest sample size is greater than 20, then a large sample approximate test statistic (say Z) is appropriate and can be computed from U , the sample sizes, and the tied value information. For large sample sizes

(Continued)

Example 1. (Continued)

with no or few ties across samples, Z approaches a normal distribution and can be tested using a standard normal table. To do this find the value in the normal table corresponding to the chosen α -value. If the computed Z is greater than the tabled value then reject the null hypothesis. If there are no or few ties, then Z is given by

$$Z = \frac{U - ((n)(m)/2)}{[((n)(m)(n + m + 1))/12]^{0.5}}$$

where U is the Mann-Whitney-Wilcoxon statistic and m and n are the sample sizes. The 0.5 superscript means to take the square root of the quantity in the square brackets.

The data in this example have three sets of ties between the two samples so this approximation is probably acceptable. The method that corrects for ties will be illustrated later in this example. Here is the uncorrected value, for this example, of Z , to three decimal places.

$$Z = \frac{493.5 - ((33)(17)/2)}{[((33)(17)(33 + 17 + 1))/12]^{0.5}}$$

$$Z = \frac{493.5 - 280.5}{(28611 / 12)^{0.5}} = \frac{213}{48.829} = 4.362$$

This computed value of Z is compared with the value found in a standard normal table for the appropriate α -level (two-tailed). For α -values of 0.10, 0.05, 0.02, 0.01, and 0.001 the standard normal table values are 1.645, 1.960, 2.326, 2.576, and 3.291. Since the computed Z is greater than 1.960 with $\alpha = 0.05$, the null hypothesis is rejected. Note that the hypothesis would also be rejected at the $\alpha = 0.001$ level as well since that tabled standard normal value = 3.291.

If there are ties, then Z is given by

$$Z = \frac{U - [(m)(n) / 2]}{\left[\left[\frac{(m)(m)}{(m + n)(m + n - 1)} \right] \left[\frac{(m + n)^3 - (m + n) - \Sigma T}{12} \right] \right]^{0.5}}$$

where ΣT is the correction factor for ties explained above. This formula looks complicated, but most of the terms are functions of the sample sizes m and n . The 0.5 superscript means to take the square root of the quantity in square brackets. For the data in this example, the corrected Z to three decimal places is

(Continued)

Example 1. (Concluded)

$$Z = \frac{493.5 - [(17)(33) / 2]}{\left[\left[\frac{(17)(33)}{(17 + 33)(17 + 33 - 1)} \right] \left[\frac{(17 + 33)^3 - (17 + 33) - 138}{12} \right] \right]^{0.5}}$$

$$Z = \frac{493.5 - 280.5}{\left[\left[\frac{561}{(50)(49)} \right] \left[\frac{125,000 - 50 - 138}{12} \right] \right]^{0.5}}$$

$$Z = \frac{213}{[(0.22897)(10,401)]^{0.5}} = 4.365$$

Since the computed Z is larger than the standard normal value for $\alpha = 0.05$ the null hypothesis is again rejected. Notice that the corrected Z value is only slightly larger than the uncorrected Z computed above. Thus the influence of ties for these data is very small.

The Kruskal-Wallis test is used to determine if one sample tends to have larger or smaller values along the habitat gradient than any other sample. The null hypothesis is that the two or more batches being compared have the same location (median). The batches being compared are pooled and the combined sample ranked. If the groups come from the same population, then a ranking that puts all of one group ahead of another group or groups is highly unlikely.

If the null hypothesis, that the batches are the same, is true, then any combination of ranks assigned to the different batches is equally probable. In the Mann-Whitney-Wilcoxon test a function of the rank sum of one of the groups is used to test whether one rank sum is excessively large or small (Table 10). In the Kruskal-Wallis test a function of all the rank sums, adjusted for group size, is used to test whether one or more rank sums, among many, is excessively large or small.

If the null hypothesis is rejected then at least one of the groups is different from the others; which group or groups are different and in what way has to be determined by other methods. For example, Figure 8 shows the depth distribution of adult white suckers as determined by three data collection methods. The Kruskal-Wallis test indicates that the methods differ statistically ($p = 0.0005$), and inspection of the box plots in Figure 8 reveals that the only difference is the greater depths found by underwater observation compared to the depths observed using each of the other methods, which in turn are not very different from each other.

Sometimes intergroup differences are not so obvious that inspection of the data can reveal them; other times you may want a statistical way to conclude which groups are different. Such tests are called multiple comparison or multiple contrast tests. Nonparametric versions can be found in Zar 1984, Sokal and Rohlf (1981), and Conover (1980).

Calculation of the Kruskal-Wallis test is illustrated in Example 2 for the focal point velocity distribution of adult white suckers as determined by three data collection methods. Box plots of these data appear in Figure 7. The null hypothesis is that there are no differences in velocities found by the three sampling techniques. The first step is to pool the data and assign ranks. Tied velocity values are assigned an average rank as illustrated in the example. The test statistic C is computed from the rank sums for each method of observation. For very small samples (three groups with less than five observations each and no ties), published tables give the exact probabilities of the test statistic (Conover 1980; Iman et al. 1975). For larger samples, the test statistic is evaluated using a chi-square table. A better approximation of the p-value for this test can be obtained by a variation of the multi-response permutation procedure, which I discuss later. That approximation uses more of the information in the data. Example 2 follows the presentations by Sokal and Rohlf (1981) and Conover (1980).

THE KOLMOGOROV-SMIRNOV TEST

Two batches of suitability curve data can be compared to see if they differ in location, shape, spread, or any other features of their distributions by the Kolmogorov-Smirnov test. This test should not be confused with the Kolmogorov one-sample goodness of fit test, which compares a single batch to the normal, chi-square, uniform, or other distribution given independently of the data. Since the Kolmogorov-Smirnov test is sensitive to any differences between the distributions of the two batches, it is more versatile than the Mann-Whitney-Wilcoxon test. However, the Kolmogorov-Smirnov test is far less sensitive to differences in location alone than the Mann-Whitney-Wilcoxon test (Conover 1980).

Generalization of the Kolmogorov-Smirnov test to compare more than two batches, or groups, is possible, but tables for more than three groups or unequal sample sizes have not been developed (Conover 1980). In practice this test must be considered a two-group test only. Also, no parametric counterpart exists for the Kolmogorov-Smirnov test.

The basis for the Kolmogorov-Smirnov two-sample test is different from the rank tests described above. This test uses the empirical cumulative distribution functions of the two batches; the test statistic is the largest difference between the two functions. The empirical cumulative distribution function of a batch gives the fraction of the data less than or equal to each individual value in the batch. This cumulative distribution function is called "empirical" because it is based on the data. Other cumulative distributions are not empirical in that they are based on some theoretical distribution such as the normal distribution. To use these distributions requires an assumption beyond the data.

Example 2. Kruskal-Wallis test for two or more groups. Data are three samples of the focal point velocity (FPV) used by adult white suckers. The samples were collected by three different methods: electrofishing, (E, sample size = $n_1 = 13$), visual observation (V, $n_2 = 17$), and underwater observation (U, $n_3 = 9$). The pooled data are arranged below by the value of the velocity variable in feet per second. The average rank and t/T columns are explained below.

Rank	FPV (f/s)	Average rank (R)	t/T	Method	Rank	FPV (f/s)	Average rank (R)	t/T	Method
1	0.0	5	9/720	E	18	0.2	22	9/720	E
2	0.0	5		E	19	0.2	22		E
3	0.0	5		E	20	0.2	22		E
4	0.0	5		E	21	0.2	22		E
5	0.0	5		V	22	0.2	22		E
6	0.0	5		V	23	0.2	22		V
7	0.0	5		V	24	0.2	22		V
8	0.0	5		V	25	0.2	22		V
9	0.0	5		U	26	0.2	22		U
10	0.1	13.5	8/504	E	27	0.3	29.5	6/210	E
11	0.1	13.5		E	28	0.3	29.5		V
12	0.1	13.5		V	29	0.3	29.5		V
13	0.1	13.5		V	30	0.3	29.5		V
14	0.1	13.5		V	31	0.3	29.5		U
15	0.1	13.5		V	32	0.3	29.5		U
16	0.1	13.5		V	33	0.4	34		V
17	0.1	13.5		U	34	0.4	34	3/24	U
					35	0.4	34		U
					36	0.5	36		E
					37	0.7	37		U
					38	0.8	38		U
					39	1.1	39		V

The hypotheses are

H_0 : The focal point velocities observed by the three methods are the same.

H_1 : At least one of the methods tends to yield larger observations than at least one of the others.

The assumptions are

- 1) All samples are random selections from their populations.
- 2) The measurement scale is at least ordinal.
- 3) If there is a difference among the populations, then at least one of them differs from another in location.

(Continued)

Example 2. (Continued)

Assign ranks.

Combine the three samples into a single list, as above, ordered by the habitat variable. Assign average ranks (R) 1 to 39 ($n_1 + n_2 + n_3$).

The average ranks for tied velocity values (shown in brackets above) are calculated by averaging the ranks that would have been assigned were there no ties. For example, the six observations with a velocity value of 0.3 feet per second have an average rank of

$$(27 + 28 + 29 + 30 + 31 + 32) / 6 = 29.5.$$

Compute test statistic.

Arrange the data by group, as below, showing the average rank for each focal point velocity value.

	Electro-shock		Visual observation		Underwater observation	
	Velocity	Rank	Velocity	Rank	Velocity	Rank
1	0.0	5	0.0	5	0.0	5
2	0.0	5	0.0	5	0.1	13.5
3	0.0	5	0.0	5	0.2	22
4	0.0	5	0.0	5	0.3	29.5
5	0.1	13.5	0.1	13.5	0.3	29.5
6	0.1	13.5	0.1	13.5	0.4	34
7	0.2	22	0.1	13.5	0.4	34
8	0.2	22	0.1	13.5	0.7	37
9	0.2	22	0.1	13.5	0.8	38
10	0.2	22	0.2	22		
11	0.2	22	0.2	22		
12	0.3	29.5	0.2	22		
13	0.5	36	0.3	29.5		
14			0.3	29.5		
15			0.3	29.5		
16			0.4	34		
17			1.1	39		

Sum of ranks $\Sigma R_1 = 222.5$

$\Sigma R_2 = 315.0$

$\Sigma R_3 = 242.5$

Sample size $n_1 = 13$

$n_2 = 17$

$n_3 = 9$

Total sample $N = 13 + 17 + 9 = 39$

(Continued)

Example 2. (Continued)

Note the sample sizes and compute the sum of ranks ΣR_i for each group. The capital Greek letter sigma (Σ) followed by R_i means to take the sum of all the ranks for sample i . Thus ΣR_3 is the sum of the 9 numbers for the underwater observation sample, i.e.,

$$\Sigma R_3 = (5 + 13.5 + 22.5 + \dots + 34 + 37 + 38) = 242.5$$

Compute the Kruskal-Wallis test statistic, H , as follows. The numbers 12 and 3 appear in every test and are not computed from the sample.

$$H = \frac{12}{N(N+1)} \left[\Sigma \frac{(\Sigma R_i)^2}{n_i} \right] - 3(N+1)$$

$$H = \frac{12}{39(40)} \left[\frac{(222.5)^2}{13} + \frac{(315)^2}{17} + \frac{(242.5)^2}{9} \right] - 3(40)$$

$$H = \frac{12}{1560} [3808.17 + 5836.76 + 6534.0] - 120$$

$$H = \frac{12}{1560} [16178.96] - 120 = 4.45$$

Correction for ties.

If there are tied velocity values the computed H is divided by a correction factor for ties. The values used for the correction factor appear in the column labeled t/T in the combined data table above. For each set of ties between or within groups, note the number of ties (t) in the set. For example $t = 5$ for the set tied with average rank equal to 5, and $t = 8$ for the set of eight values tied for rank 13.5. Now for each set of ties compute a correction factor $T = (t^3 - t)$, which for easier computation can be rewritten as $T = (t-1)t(t+1)$. For the set of ties of average rank 5 in this example $T = (9^3 - 9) = 720$ or

$$T = (9-1)9(9+1) = (8)(9)(10) = 720.$$

The correction factor for ties uses the sum of all the T values (ΣT). In this example $\Sigma T = (720 + 504 + 720 + 210 + 24) = 2178$. The correction factor, C , is computed as follows.

(Continued)

Example 2. (Concluded)

$$C = 1 - \left[\frac{\Sigma T}{(N - 1)(N)(N + 1)} \right]$$

$$C = 1 - \left[\frac{2178}{(39 - 1)(39)(39 + 1)} \right]$$

$$C = 1 - \left[\frac{2178}{(38)(39)(40)} \right]$$

$$C = 1 - 0.03674 = 0.96$$

Compute the corrected Kruskal-Wallis test Statistic, H_c , by division.

$$H_c = H / C = 4.45 / 0.96 = 4.62.$$

Test significance of H or H_c .

For three samples each with five observations or fewer and no ties the exact p-value of H can be found in Zar (1984) and Conover (1980). Conover (1980) also cites more extensive exact tables in Iman, Quade, and Alexander (1975). Sokal and Rohlf (1981) refer to exact tables in Kruskal and Wallis (1952).

For larger samples the test statistic, H or H_c , is distributed approximately as chi-square with degrees of freedom one less than the number of samples (groups) being compared, i.e., $3 - 1 = 2$ for the current example. Consult a chi-square table and find the tabled number corresponding to the degrees of freedom and the selected α -value. For 2 degrees of freedom and $\alpha = 0.05$ the tabled chi-square value is 5.99. Since the computed $H_c = 4.62$ is less than 5.99, the null hypothesis of no difference among samples cannot be rejected. Notice that since the chi-square value for a confidence level of 0.10 is 4.605, the p-value in this case is just less than 0.10.

Figure 10 shows day and night velocity data for white sucker juveniles in two ways. The top of the figure gives the familiar frequency plots or bar graphs of number of observations for each velocity value. The bottom of the figure shows the empirical cumulative distribution functions for each batch. Here, for each velocity value, the height of the empirical function is the sum of the observations for that velocity value and for all the velocity values below it. This sum is expressed as a fraction or proportion of the number of observations in the batch; therefore, the vertical axis ranges from 0.0 to 1.0 and is the same for all batches.

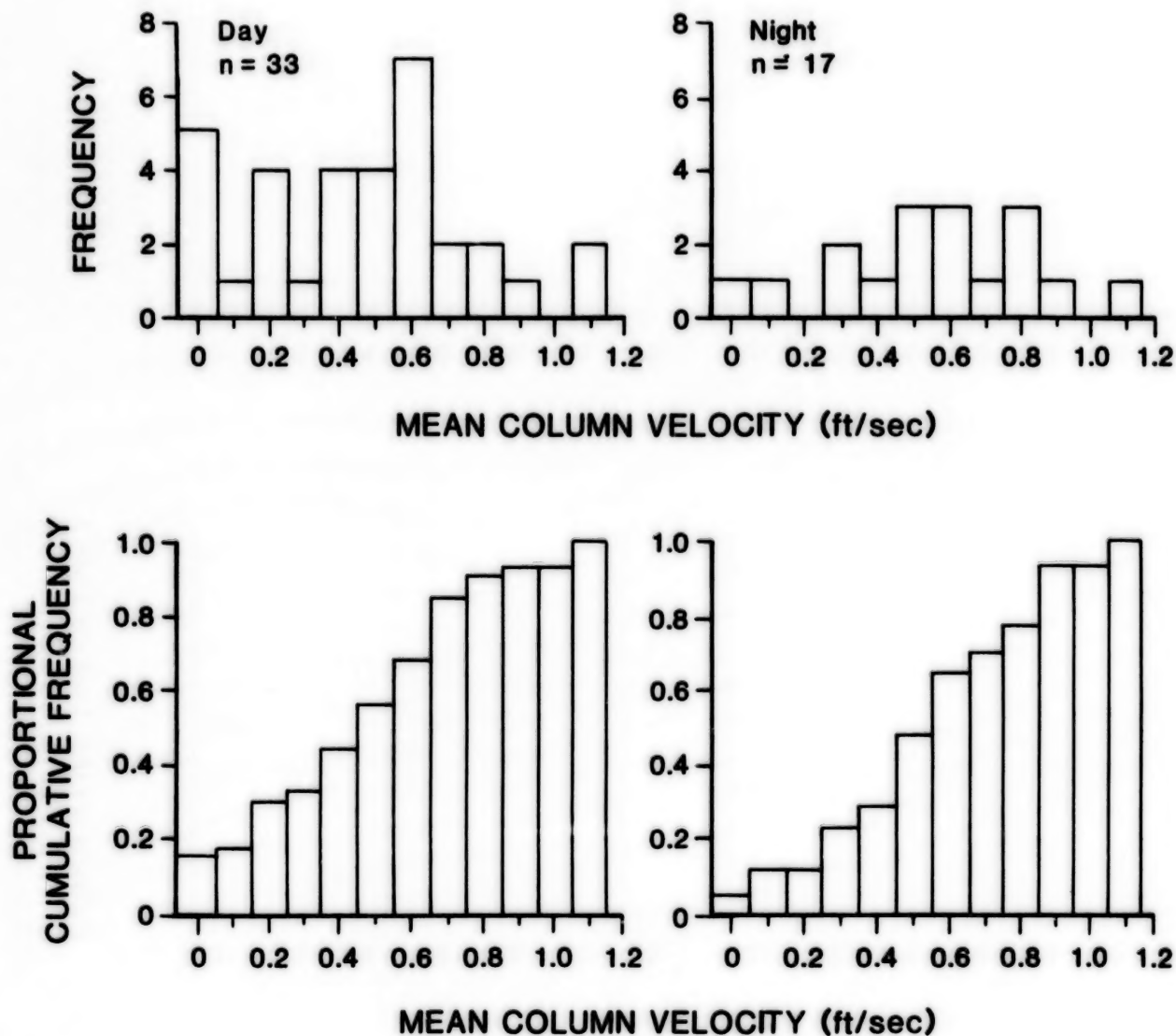


Figure 10. Frequency (top) and empirical cumulative frequency distributions (bottom) of mean column velocity used in the daytime and nighttime by juvenile white suckers.

To compare two batches it is convenient to plot and connect only the tops of the bars of the empirical cumulative distribution function, making a stair step graph for each batch. These can be plotted together, since they have the same scales. Figure 11 shows such plots for the day and night velocity values given in Figure 10.

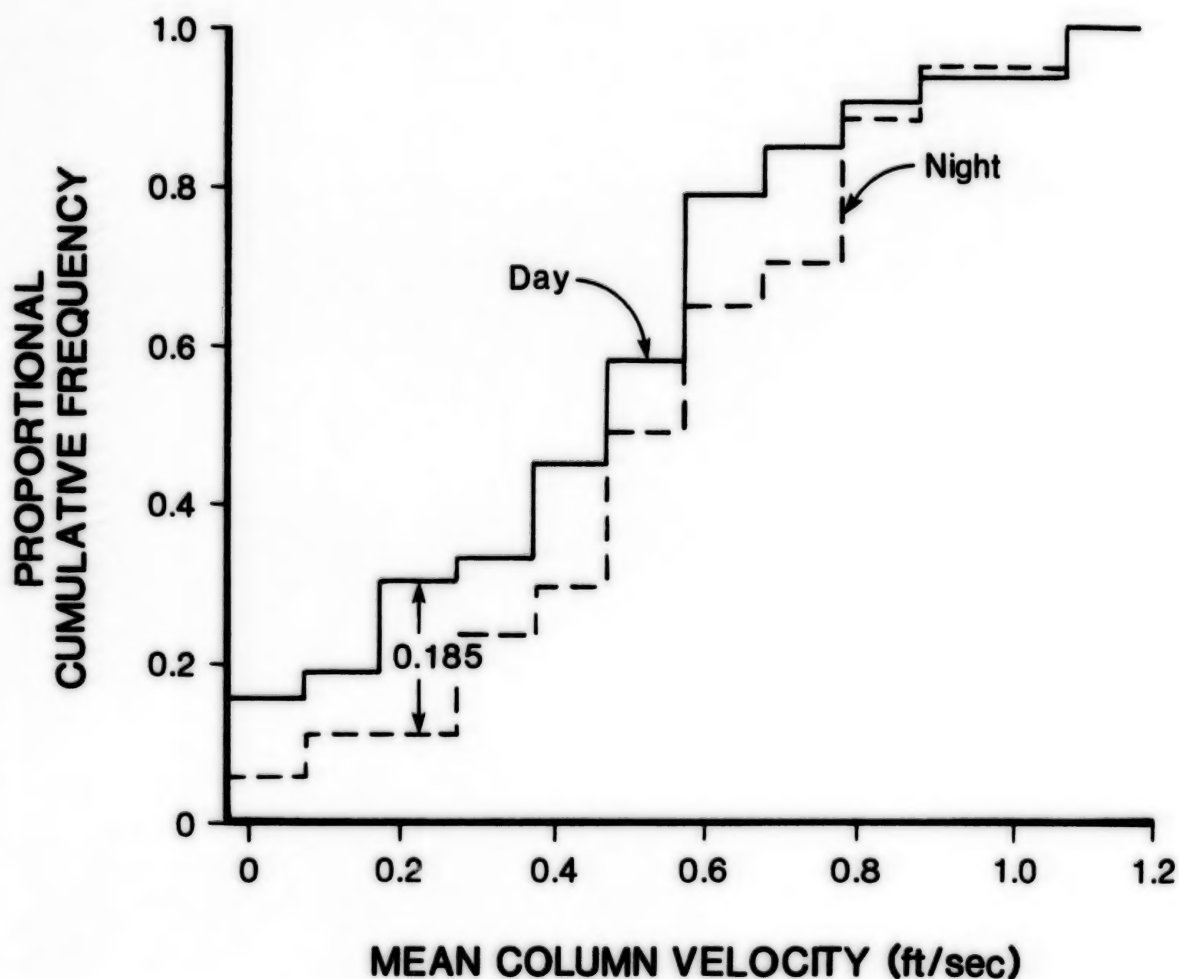


Figure 11. Empirical cumulative frequency distributions of daytime and night-time mean column velocity used by juvenile white suckers. The maximum vertical distance between the distributions is indicated.

The essence of the Kolmogorov-Smirnov two-sample test is that any differences between the distributions of the two batches will show up as differences in the empirical cumulative distributions. Thus, the greatest vertical discrepancy between the two cumulative distributions is a measure of the difference between the two batches. The test statistic, D , for the two-tailed test just is the greatest vertical distance between the two curves.

The two batches of velocity data for juvenile white suckers (Figures 10 and 11) appear to be similar. They have the same range and about the same median, but note that the daytime batch tends to have relatively more observations from 0.0 to 0.6 feet per second. This difference between the two

batches can be seen in the empirical cumulative distributions in that the daytime curve is generally above the nighttime curve and that the largest vertical distance occurs between 0.2 and 0.3 feet per second.

The null hypothesis is that the two batches of data are from the same distribution. The null hypothesis is rejected if the maximum distance, D , is large compared to the distances expected if the null hypothesis is true. The rationale of the Kolmogorov-Smirnov test is explained as follows. If the null hypothesis is true, then the two samples, with sample sizes m and n , can be considered as two samples of the same underlying distribution. Combine the data and select all the possible combinations of two samples with sizes m and n and compute the D statistic. Now each of the possible combinations is equally probable and will include the two samples actually obtained. If the possible D values are ordered from smallest to largest, counting duplicates, then the proportion of D 's greater than or equal to the D actually computed is the p -value of the observed D . If this p -value is smaller than the chosen value of α (e.g., 0.05), then the null hypothesis is rejected. Of course, this process of finding all the possible subsets of sizes m and n and determining all the possible D values need not be repeated when doing a Kolmogorov-Smirnov test. Rather, statisticians have already done all that work and produced tables of D values corresponding to different sample sizes and α -levels.

Calculation of the Kolmogorov-Smirnov test is illustrated in Example 3 for the day and night velocity distribution of juvenile white suckers already shown in Figures 10 and 11. The null hypothesis is that there are no day/night differences in velocities used.

The data used in Example 3 are the companion data to the white sucker depth data used in Example 1 to demonstrate the Mann-Whitney-Wilcoxon test. If the depth data are subjected to the Kolmogorov-Smirnov test, daytime and nighttime depths are found to differ significantly ($D = 0.588$, p -value < 0.05), which corresponds to the conclusion drawn from the Mann-Whitney-Wilcoxon test. These tests will often, but not always, give similar results for the same set of data. For comparing two sets of suitability index curve data, either test can be used. Choose the Mann-Whitney-Wilcoxon test to test for differences in the location of the curves; choose the Kolmogorov-Smirnov test to test for any differences in the location, spread, or shape of the curves. Example 3 follows the presentation by Conover (1980).

MULTI-RESPONSE PERMUTATION PROCEDURES (MRPP)

The multi-response permutation procedures (MRPP) are a fairly recent set of methods that depend strictly on the observed data (Mielke et al. 1976; Mielke et al. 1981a,b; Mielke 1984, 1986). These techniques are not yet described in textbooks, but have been successfully applied in meteorology (see references cited in Mielke 1986), archeology (Berry et al. 1983), public health (Mielke et al. 1983), and biology (Zimmerman et al. 1985; Biondini et al. 1985). These procedures are powerful, make efficient use of small samples, and can viably replace many standard parametric and nonparametric tests. But

Example 3. Kolmogorov-Smirnov test. Data are two samples of the mean column velocity used by juvenile white suckers (Figure 10). One set is for daytime use (sample size = $m = 33$), the other is for nighttime use ($n = 17$). The data for each group are arranged below by the value of the velocity variable. The frequency columns give the number of observations at each velocity value, the other columns are explained below.

Velocity feet/sec	Frequency		Cumulative frequency		Cumulative rel. freq.		Absolute value of difference
	Day	Night	Day	Night	Day	Night	
0.0	5	1	5	1	.152	.059	.093
0.1	1	4	6	2	.182	.118	.064
0.2	4	0	10	2	.303	.118	.185 *
0.3	1	2	11	4	.333	.235	.098
0.4	4	1	15	5	.454	.294	.160
0.5	4	3	19	8	.576	.488	.079
0.6	7	3	26	11	.788	.647	.141
0.7	2	1	28	12	.848	.706	.142
0.8	2	3	30	15	.909	.882	.027
0.9	1	1	31	16	.939	.941	.002
1.0	0	0	31	16	.939	.941	.002
1.1	2	1	33	17	1.000	1.000	.000
Total	33	17					

* largest value

The hypotheses are

- H_0 : The day and night velocities are the same.
 H_1 : The day and night velocities are different.

The assumptions are

- 1) Both samples are random selections from their populations.
- 2) The measurement scale is at least ordinal.

Compute the empirical cumulative frequency distribution.

The cumulative frequency column for the daytime data in the above table is a running total of the daytime frequencies. Thus the third entry is $5 + 1 + 4 = 10$ or the sum of the number of daytime observations for velocities 0.2 feet per second or less. The nighttime cumulative distribution is computed similarly; the third entry is $1 + 1 + 0 = 2$. To compute the cumulative relative frequency each cumulative frequency is expressed as a fraction of the sample size. Thus the first number in the daytime column is $5/33 = 0.152$ (to three decimals), and the last number is $33/33 = 1.0$. Similarly, for the nighttime data, the third entry down the cumulative relative frequency column is $2/17 = 0.118$, and the fourth is $4/17 = 0.235$.

(Continued)

Example 3. (Continued)

Compute the test statistic.

The last column in the above table is the absolute value of the difference between the day and night cumulative relative frequencies for each velocity value. Thus the second value is $|0.182 - 0.118| = 0.064$, and the third is $|0.303 - 0.118| = 0.185$, where the vertical bars indicate the absolute value of the difference. Inspection of this column reveals that 0.185 is the largest value, and this becomes the test statistic, D . Notice that all the computations given in the above table may not need to be computed, for often inspection of the data or the plots of the empirical cumulative distribution functions (Figure 11) will reveal the region where the greatest difference between the two functions occurs.

Test significance of D .

For sample sizes 20 or less, the computed test statistic D can be compared to critical values found in published tables for selected values of α (Conover 1980). Rohlf and Sokal (1981) present tables where the number to be compared with the tabled values is not D but rather D times the product of the sample sizes (e.g., $(33)(17)(0.185) = 103.78$ in the present case). If the computed test statistic D is smaller than the tabled number, then the null hypothesis cannot be rejected.

For large sample sizes the computed test statistic can be compared to an approximate critical value D_α , computed as follows.

$$D_\alpha = F_\alpha \left[\frac{m+n}{mn} \right]^{0.5}$$

Where m and n are the sample sizes, the superscript 0.5 means to take the square root of the quantity in brackets, and F_α is a factor selected from the table below corresponding to the α -level for the test.

α -level	0.20	0.10	0.05	0.02	0.01
F_α	1.07	1.22	1.36	1.52	1.63

These factors are for the two-tailed tests at each α -level.

For the current example and $\alpha = 0.05$, the approximate critical value is

$$D_{.05} = 1.36 \left[\frac{33+17}{33(17)} \right]^{0.5}$$

$$D_{.05} = 1.36 \left[\frac{50}{561} \right]^{0.5}$$

(Continued)

Example 3. (Concluded)

$$D_{.05} = 1.36 [0.089]^{0.5} = 0.406$$

Since the computed test statistic $D = 0.185$ is less than 0.406, the null hypothesis is not rejected. There is no evidence that the daytime and nighttime use of velocity by these fish differs.

the strength of these methods is their intuitive basis, with its natural geometric interpretation.

MRPP can be used to compare two or more batches of habitat suitability data. The test can thus be used in place of the Mann-Whitney-Wilcoxon and Kolmogorov-Smirnov two-sample tests and the Kruskal-Wallis test for two or more groups. The null hypothesis is that the samples have the same distribution along the habitat variable. The alternative hypothesis is that the values for at least one sample are clustered or concentrated on one part of the habitat variable compared to the others. Samples with different medians, shapes, or spread can show such clustering along the habitat gradient. For MRPP the data must be random and at least ordinal, or for categorical data the variables must be dichotomous. The analysis can be performed on ranked or unranked data, but this may lead to different results.

MRPP may take many forms. For example, changing the details of the computation in various ways gives procedures that mimic a surprising range of parametric and nonparametric tests, including the parametric analysis of variance, the two-sample and paired-sample t-tests, and the Mann-Whitney-Wilcoxon, Kruskal-Wallis, and Wald-Wolfowitz runs test (Mielke 1986). I will present a relatively simple version of MRPP relating the response of populations to a single habitat variable. But the more general technique is truly multi-response or multivariate and can be used to detect differences on many habitat variables at once, in which case MRPP can mimic such tests as Hotelling's T^2 and multivariate analysis of variance. Indeed, the technique may be understood more easily with a bivariate example than with the univariate example given below. Examples using the response to two variables are given in Mielke (1981), Biondini et al. (1985), and Zimmerman et al. (1985).

One practical problem with MRPP is that it cannot be computed by hand for even modest sample sizes. The computational algorithm and the FORTRAN computer code for MRPP are available in print (Berry and Mielke 1983). A version of MRPP that runs on IBM PC, XT, and AT compatible computers is available from the National Ecology Research Center of the U. S. Fish and Wildlife Service. It is located on the National Ecology Research Center (NERC) electronic bulletin board, which can be reached by 303-226-9365 (or FTS 323-5365). Transmission rates are 300, 1200, and 2400 bits per second; word size is eight bits with no parity bit and one stop bit. Help on using the bulletin board can be obtained at 303-226-9335 (or FTS 323-5335). After contacting the

bulletin board (logging on), look in the bulletins section for information on how to get (download) the MRPP programs. The executable program, FORTRAN 77 source code, documentation, and sample data are available. The FORTRAN 77 code is generic and can be compiled to run on other micro-, mini-, and main-frame computers.

I will explain the basis of MRPP by comparing two batches of artificial data, one with three and one with four observations. I will give the results for some actual curve data, but cannot illustrate those computations.

Two groups of observations for values of an arbitrary habitat gradient are shown in Figure 12. Each observation of a fish is identified in the frequency plot by its group (A or B) and observation number (A_1 - A_3 or B_1 - B_4). This is the type of data often used for habitat suitability curves. The sample sizes are small to facilitate the computations. The groups might represent observations made on one life stage in two different streams or during different seasons, for example. If just two of the observations, B_1 and A_3 , were switched, then the two groups would seem to be concentrated on opposite ends of the habitat gradient. But do the observed data support the conclusion that the two groups respond differently to the habitat gradient? One way to determine this is to compare the average distance separating members of group A and the average distance separating members of group B with the intragroup distances for all the other possible ways of arranging the 7 observations into groups of 3 and 4 members.

To do this, first calculate the distance between all possible pairs of observations. The number of combinations of 7 things taken 2 at a time is $7!/((2!(7-2)!)) = 7!/(2!(5!)) = 6(7)/2 = 21$, where 7! (seven factorial) means to multiply the integers from 1 to 7 together ($1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7$). These 21 distances are given in Table 11 and a few of them are shown graphically in Figure 12. The distance, for example, of observation A_2 to B_1 is 1 unit on the horizontal axis of Figure 12, the distance from B_1 to B_3 is 3 units, and the distance between A_1 and A_2 is 0 units.

The intragroup distances are computed next. The distances between members of group A are 0, 2, and 2 and between members of group B are 3, 3, 5, 0, 2, and 2. The average distance within group A is thus 1.33 and within group B is 2.50. The average group distances are usually symbolized by the lower case Greek letter zeta (ζ). Small average group distances indicate the group is concentrated; large distances indicate the group is spread out along the habitat variable.

Now the approach of MRPP is to compare the observed average distances with those for all the possible ways these 7 observations might have formed two groups of sizes 3 and 4. If the observed distances are small among the set of possible distances, then clustering is indicated. Seven observations can be assigned to two groups with 3 and 4 members in $7!/(3!(4!)) = 35$ possible ways. The intragroup distances for a few of the 35 permutations are given in Table 12. All the permutations and the average distances are shown in Table 13.

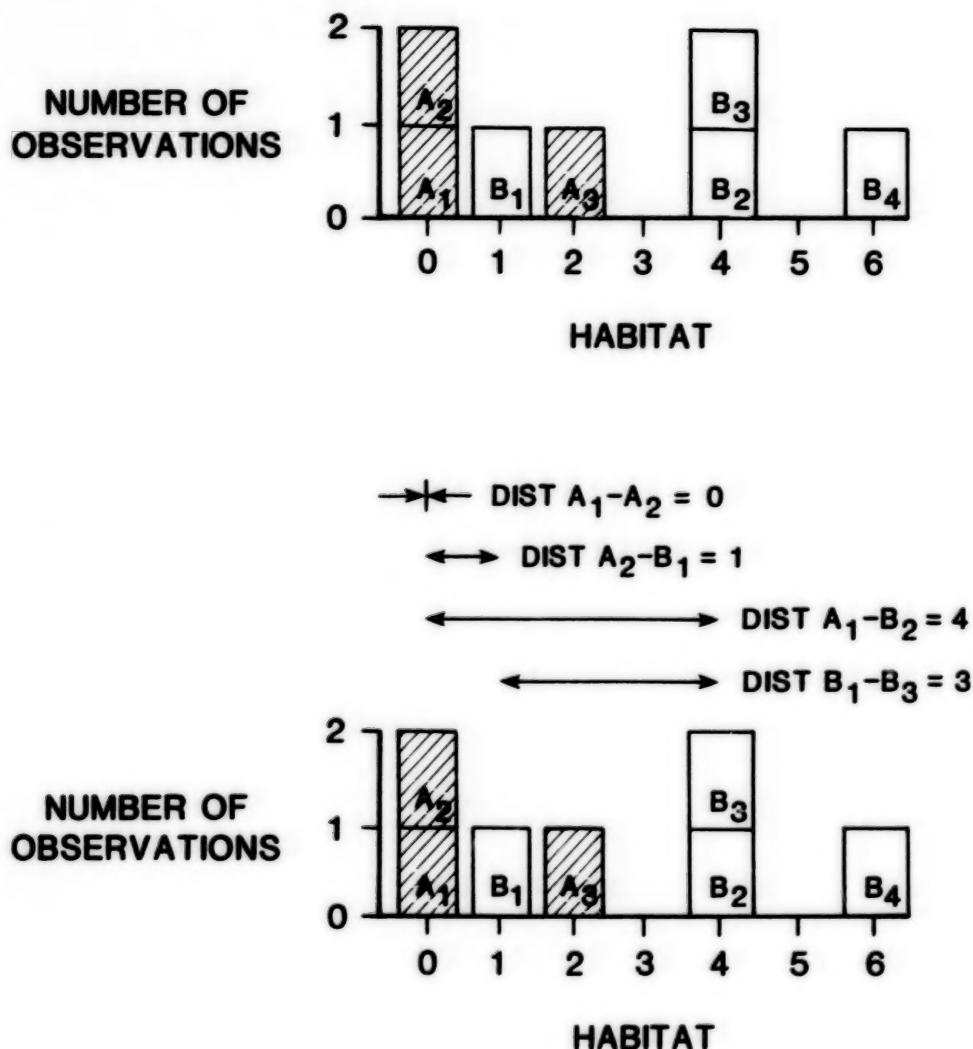


Figure 12. Sample data for demonstration of multi-response permutation procedures. The habitat values for seven observations (in two groups, A and B) are shown in the top plot. The distance between selected observations are shown in the bottom plot.

The test statistic for MRPP is the average of the observed intragroup distances weighted by group size, 3 and 4 in this case. It is usually symbolized by the lower case Greek letter delta (δ). The average intragroup distances for group A and B are 1.33 and 2.50; thus their weighted average is $[3(1.33) + 4(2.50)]/7 = 2.0 = \delta = \text{delta}$. The observed delta corresponds to permutation 1 in Table 13. To determine if the observed delta is small enough to indicate clustering it is compared to all the possible deltas this set of data could have exemplified. Inspection of the delta values in Table 13 shows

Table 11. Distances between all possible pairs of observations in Figure 12. Distances are in the units of the habitat axis.

	Pair	Distance		Pair	Distance
1	A_1-A_2	0	12	A_3-B_1	1
2	A_1-A_3	2	13	A_3-B_2	2
3	A_1-B_1	1	14	A_3-B_3	2
4	A_1-B_2	4	15	A_3-B_4	4
5	A_1-B_3	4	16	B_1-B_2	3
6	A_1-B_4	6	17	B_1-B_3	3
7	A_2-A_3	2	18	B_1-B_4	5
8	A_2-B_1	1	19	B_2-B_3	0
9	A_2-B_2	4	20	B_2-B_4	2
10	A_2-B_3	4	21	B_3-B_4	2
11	A_2-B_4	6			

that only 2 are smaller, 1.429 and 1.238, corresponding to permutation numbers 2 and 35. Since under the null hypothesis each of the 35 possible group assignments is equally probable, each has a $1/35 = 0.029$ chance of occurring. Thus, getting a delta as small or smaller than 2.0 has a probability of $3/35 = 0.086$ because 3 of the values are equal to or smaller than 2.0, the delta actually observed. The null hypothesis could not be rejected at the 0.05 α -level, but could be rejected at the 0.10 level.

To further understand this test, notice that the smallest delta, found for permutation 35 in Table 13, results for the case where the four left-most observations of Figure 12 are in one group and the three right-most in the other. If such observations had been made, then the p-value for the resulting delta would have been the smallest one possible, i.e., $1/35 = 0.029$. This potential grouping is shown in Figure 13a. The next smallest delta value, permutation 2 in Table 12, corresponds to the case where the three left-most observations are in one group and the four right-most in the other

Table 12. All pairs of distances within each group for the first five possible permutations into two groups of sizes 3 and 4 of the observations shown in Figure 12.

Permutation	Group 1	Distances	Group 2	Distances
1	A ₁ , A ₂ , A ₃	0 2 2	B ₁ , B ₂ , B ₃ , B ₄	3 3 5 0 2 2
2	A ₁ , A ₂ , B ₁	0 1 1	A ₃ , B ₂ , B ₃ , B ₄	2 2 4 0 2 2
3	A ₁ , A ₂ , B ₂	0 4 4	A ₃ , B ₁ , B ₃ , B ₄	1 2 4 0 2 2
4	A ₁ , A ₂ , B ₃	0 4 4	A ₃ , B ₁ , B ₂ , B ₄	1 2 4 3 5 2
5	A ₁ , A ₂ , B ₄	0 6 6	A ₃ , B ₁ , B ₂ , B ₃	1 2 2 3 3 0

(Figure 13b). If such observations had been made, then the p-value for the resulting delta would have been $2/35 = 0.057$. This possible result shows less clustering, since both the three and four member groups are spread out more than those in permutation 35. The third smallest delta, permutation 1, represents the sample as it was actually taken (Figure 13c). All other permutations show less clustering and accordingly have larger deltas and associated p-values.

The reason for the difficulty of computing MRPP by hand is the large number of distances and permutations that exist for samples of relatively modest size. For the 50 observations in the white sucker data given in Example 1, 1,225 distances must be calculated. A more usual data set with 100 observations per group would require that 19,900 distances be computed. The number of permutations is even larger. For the white sucker data, with two groups of 17 and 33 members, the number of permutations is $50!/(17!(33!)) = 9.847 \times 10^{10}$. This is more than the number of seconds since the year 1000 B.C. The number of permutations for two groups of 100 is $200!/((100!(100!)) = 9.05 \times 10^{58}$. This number is so large it is hard to find a physical interpretation. Try computing the number of milliseconds since the end of the Precambrian 600 million years ago.

With such large numbers, MRPP cannot be computed directly on any computer. Instead, an approximation is used that is fairly adequate when the number of permutations is as small as 1,000 (Mielke and Berry 1982). Even for the extremely small sample sizes, 3 and 4, used in the example above (with only 35 permutations), the approximate p-value of 0.069 is surprisingly close to the exact p-value of 0.086. If the total number of permutations is smaller than about 100,000 the exact MRPP probability can be computed (depending on the computer).

Table 13. All possible permutations of the observations A_1 - A_3 and B_1 - B_4 shown in Figure 12. For each permutation the average intragroup distances, delta (δ), and probability (p) for a delta as small or smaller is given. The individual distances for the first 5 permutations are shown in Table 12. Computations are described in the text.

Permutation number	Group 1 members	Group 2 members	Average dist.		Delta δ	p
			Group 1	Group 2		
1	A_1, A_2, A_3	B_1, B_2, B_3, B_4	1.333	2.500	2.000	0.086
2	A_1, A_2, B_1	A_3, B_2, B_3, B_4	0.667	2.000	1.429	0.057
3	A_1, A_2, B_2	A_3, B_1, B_3, B_4	2.667	2.833	2.762	0.371
4	A_1, A_2, B_3	A_3, B_1, B_2, B_4	2.667	2.833	2.762	0.371
5	A_1, A_2, B_4	A_3, B_1, B_2, B_3	4.000	1.833	2.762	0.371
6	A_1, A_3, B_1	A_2, B_2, B_3, B_4	1.333	4.000	2.857	0.400
7	A_1, A_3, B_2	A_2, B_1, B_3, B_4	2.667	3.500	3.143	1.000
8	A_1, A_3, B_3	A_2, B_1, B_2, B_4	2.667	3.500	3.143	1.000
9	A_1, A_3, B_4	A_2, B_1, B_2, B_3	4.000	2.500	3.143	1.000
10	A_2, A_3, B_1	A_1, B_2, B_3, B_4	1.333	3.000	2.286	0.143
11	A_2, A_3, B_2	A_1, B_1, B_3, B_4	2.667	3.500	3.143	1.000
12	A_2, A_3, B_3	A_1, B_1, B_2, B_4	2.667	3.500	3.143	1.000
13	A_2, A_3, B_4	A_1, B_1, B_2, B_3	4.000	2.500	3.143	1.000
14	A_1, B_1, B_3	A_2, A_3, B_3, B_4	2.667	3.333	3.048	0.829
15	A_1, B_1, B_3	A_2, A_3, B_2, B_4	2.667	3.333	3.048	0.829
16	A_1, B_1, B_4	A_2, A_3, B_2, B_3	4.000	2.333	3.048	0.829
17	A_1, B_2, B_3	A_2, A_3, B_1, B_4	2.667	3.167	2.952	0.657
18	A_1, B_2, B_4	A_2, A_3, B_1, B_3	4.000	2.167	2.952	0.657

(Continued)

Table 13. (Concluded)

Permutation number	Group 1 members	Group 2 members	Average dist.		Delta δ	p
			Group 1	Group 2		
19	A ₁ , B ₃ , B ₄	A ₂ , A ₃ , B ₁ , B ₂	4.000	2.167	2.952	0.657
20	A ₂ , B ₁ , B ₂	A ₁ , A ₃ , B ₃ , B ₄	2.667	3.333	3.048	0.829
21	A ₂ , B ₁ , B ₃	A ₁ , A ₃ , B ₂ , B ₄	2.667	3.333	3.048	0.829
22	A ₂ , B ₁ , B ₄	A ₁ , A ₃ , B ₂ , B ₃	4.000	2.333	3.048	0.829
23	A ₂ , B ₂ , B ₃	A ₁ , A ₃ , B ₁ , B ₄	2.667	3.167	2.952	0.657
24	A ₂ , B ₂ , B ₄	A ₁ , A ₃ , B ₁ , B ₃	4.000	2.167	2.952	0.657
25	A ₂ , B ₃ , B ₄	A ₁ , A ₃ , B ₁ , B ₂	4.000	2.167	2.952	0.657
26	A ₃ , B ₁ , B ₂	A ₁ , A ₂ , B ₃ , B ₄	2.000	3.667	2.952	0.657
27	A ₃ , B ₁ , B ₃	A ₁ , A ₂ , B ₂ , B ₄	2.000	3.667	2.952	0.657
28	A ₃ , B ₁ , B ₄	A ₁ , A ₂ , B ₂ , B ₃	3.333	2.667	2.952	0.657
29	A ₃ , B ₂ , B ₃	A ₁ , A ₂ , B ₁ , B ₄	1.333	2.833	2.190	0.114
30	A ₃ , B ₂ , B ₄	A ₁ , A ₂ , B ₁ , B ₃	2.667	2.167	2.381	0.200
31	A ₃ , B ₃ , B ₄	A ₁ , A ₂ , B ₁ , B ₂	2.667	2.167	2.381	0.200
32	B ₁ , B ₂ , B ₃	A ₁ , A ₂ , A ₃ , B ₄	2.000	3.167	2.667	0.229
33	B ₁ , B ₂ , B ₄	A ₁ , A ₂ , A ₃ , B ₃	3.333	2.333	2.762	0.371
34	B ₁ , B ₃ , B ₄	A ₁ , A ₂ , A ₃ , B ₂	3.333	2.333	2.762	0.371
35	B ₂ , B ₃ , B ₄	A ₁ , A ₂ , A ₃ , B ₁	1.333	1.167	1.238	0.029

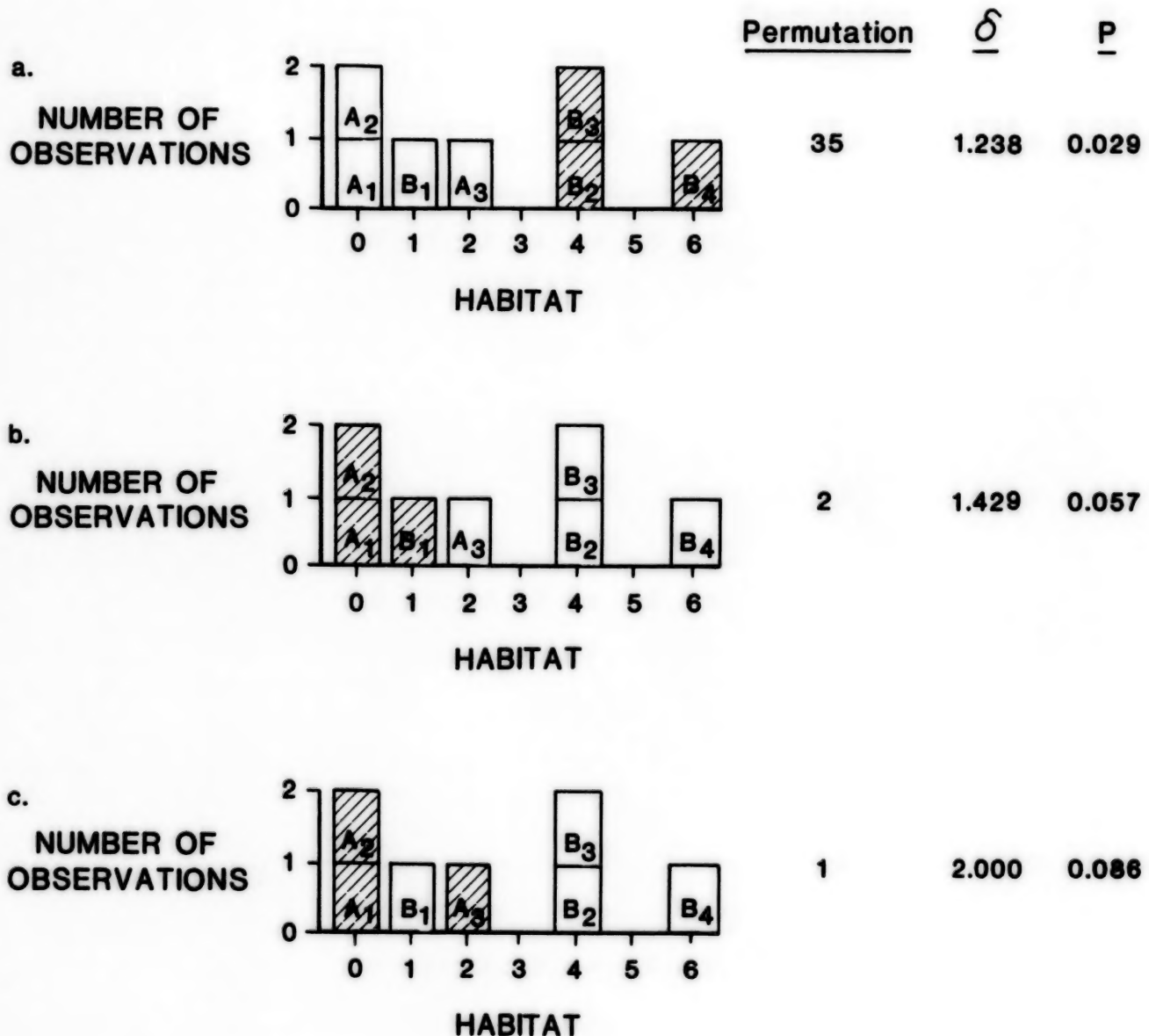


Figure 13. Two of the possible (a, b) and the observed (c) permutations of Table 13 for the observations shown in Figure 12. The delta (δ) and p-values are explained in the text.

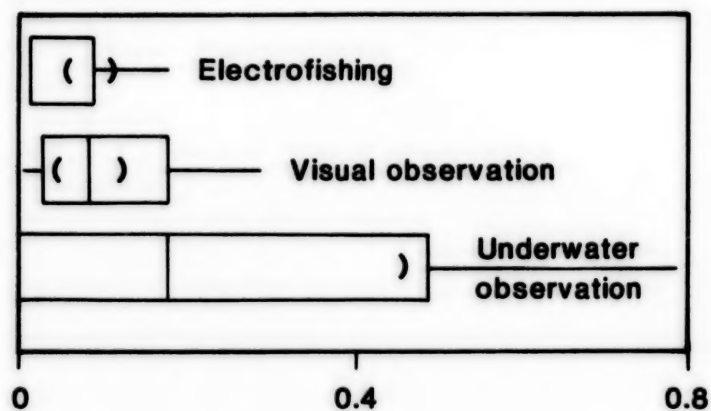
The results of MRPP tests for real data are shown in Table 14, along with the results obtained by other nonparametric methods. Included are results for the data examined in Examples 1-3 and for the data shown in Figure 7. The same conclusions about rejecting the null hypothesis would be drawn in most cases. But for mean column and focal point velocity suitability for juvenile white suckers as measured by three different methods, the conclusions drawn

Table 14. Comparison of statistical analyses by four methods on several sets of data. The statistical tests are the Mann-Whitney-Wilcoxon two-sample test (M-W-W), Kolmogorov-Smirnov two-sample test (K-S), Kruskal-Wallis analysis of variance (K-W), and the multi-response permutation procedure (MRPP). N/A means that the test is not applicable to the data. The entries in the table are p-values.

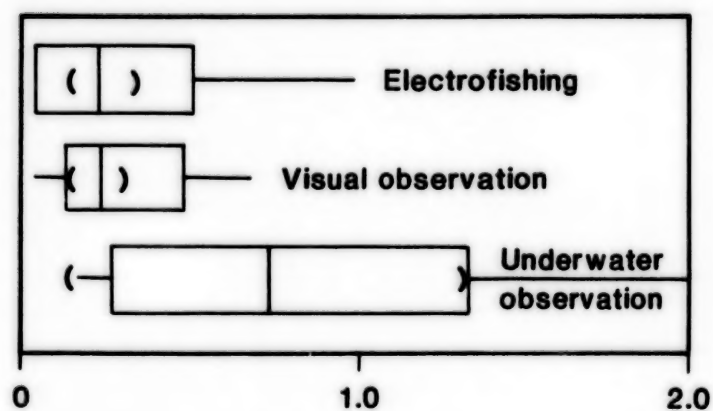
Data	M-W-W	K-S	K-W	MRPP
Juvenile white sucker day versus night				
Depth	0.000	0.001	N/A	0.000002
Mean column velocity	0.244	0.835	N/A	0.402
Focal point velocity	0.110	0.564	N/A	0.191
Juvenile white sucker three methods				
Depth	N/A	N/A	0.0006	0.00007
Mean column velocity	N/A	N/A	0.0355	0.0047
Focal point velocity	N/A	N/A	0.0869	0.0191
Adult white sucker three methods				
Depth	N/A	N/A	0.0005	0.00011
Mean column velocity	N/A	N/A	0.6187	0.8782
Focal point velocity	N/A	N/A	0.0991	0.1172

from the MRPP and Kruskal-Wallis tests may differ depending on the α -level chosen. The MRPP detects and the Kruskal-Wallis test does not detect differences in mean column velocity at the 0.01 level. Similar conclusions would follow for focal point velocity at the 0.05 level. For either velocity measurement, the MRPP assigns a probability about an order of magnitude lower than the Kruskal-Wallis test.

These results merit scrutiny. The box plots for these data are shown in Figure 14. Underwater observations of these fish result in a broad range of mean column and focal point velocities compared to the other two methods of observations. But notice that the 95% confidence intervals around the medians for underwater observation encompasses the medians of the other two in each group of box plots. Thus the Kruskal-Wallis test tends to find little or no difference among these methods, since it is a mean or location based test. The MRPP detects a significant clustering or grouping in these batches, indicating differing sensitivity to habitat by these three observational



FOCAL POINT VELOCITY (ft/sec)



MEAN COLUMN VELOCITY (ft/sec)

Figure 14. Box plots for focal point and mean column velocity for juvenile white suckers as determined by three methods of observation.

methods. The electrofishing and visual observations are clumped at the low end of the velocity gradient, whereas underwater observations are spread out. Now for the purposes of comparing habitat suitability curves, or in this case methods, the results of MRPP are more agreeable, for it is the shape and spread of the curve data that are important, more important than the position of a single summary value such as the median or mean.

This example shows the importance of specifying exactly what differences among curve data to test for. While you should carefully choose the test that will discriminate the differences in question, you should refrain from computing several different kinds of tests and then selecting the one to present after inspecting all the p-values.

The last example of MRPP I will mention concerns categorical data. Unlike the other tests discussed in this paper, MRPP can be used on categorical data so long as the variable is at least dichotomous. The presence and absence data for rock cover used by adult white suckers was presented earlier (Table 8). If presence and absence are scored 1 and 0, then the MRPP results show no difference for the proportion of fish using rock cover detected by the three data collection methods. The p-value is 0.137.

Since MRPP is such a versatile yet relatively unknown technique, it is worth describing some of its variations, especially those showing its relation to other commonly used statistical procedures. In the example described above, the distances shown in Figure 12 are ordinary Euclidean distances, the distances one would measure directly with a ruler, for example, or would compute by subtracting coordinate values. But, other ways to measure this distance are presupposed by other statistical methods. For example, the t-test, one-way analysis of variance, and the rank tests described in this paper replace the ordinary Euclidean distance with the square of Euclidean distance. If MRPP is computed using squared distances and a slightly different weighting factor for the intragroup average distances, then the t and F statistics can be written as a function of delta. Further, if the analysis is performed on ranked data, then the test statistics of the Mann-Whitney-Wilcoxon and Kruskal-Wallis tests are a function of delta (Mielke 1984, 1986). Also, the approximation used in MRPP to estimate p-values for these tests is more accurate than the standard approximations. This is because MRPP uses information in the data about the shape of the distribution of the test statistic.

One reason for pointing out that these four classic tests (and others) can be mimicked by MRPP is to emphasize an important advantage of MRPP. If the distance measure, or geometry, implied in the analysis matches the geometry of the data, then the analysis space and the data space are said to be congruent. Or put another way, the analysis will be carried out in a way commensurate with the ordinary intuitive grasp an investigator has of the data. Using ordinary Euclidean distance in MRPP insures this congruence.

Consider for example how outliers or extreme values often dominate parametric tests. The reason for this is that, being outliers, they have large distances to the rest of the values, and an analysis that squares these distances distorts them even more. As distances (greater than 1.0) get bigger their squares get bigger in disproportion. Even data without outliers and ranked data will be distorted in the analysis when squared distances are used.

Squared distance also violates one of our intuitions about distance measures, namely that the distance between two objects cannot be larger than the sum of the distances from each of these objects to another (this is the triangular inequality axiom for true distance measures). For example, 3

objects on a straight line might occupy positions 0, 2, and 3, then the distance between the first and third equals 3, which equals the sum of 2 and 1. However, the squared distance, 9, is larger than the sum of squares of 2 and 1, that sum being 5. In viewing frequency plots of habitat suitability curves we do not naturally, and probably never should, think in terms of squared distances along the habitat gradient.

It also becomes difficult to understand statistical tests that so construe geometric relations. It's not that methods based on squared distance are invalid, rather that the behavior of such a statistical test, especially in the critical regions, becomes obscure relative to the investigators perception of the data. Other more technical problems that arise when the analysis and data space are not congruent are presented in Mielke (1984, 1986).

The MRPP does not oblige us to analyze data in a way conflicting with our understanding of the data. Rather, we can select the appropriate analysis given our knowledge of the data and our purposes. For some analyses, emphasis on outliers may well be important, and squared, even cubed, Euclidean distance may be the appropriate choice. It is best to perform a test where the analysis space is congruent with the data space and the purposes of investigation. MRPP gives this choice.

SUMMARY

The purpose of this paper is to describe a variety of graphic and statistical methods for exploring and comparing habitat suitability curve data. Since many things have to be considered in preparing and analyzing data, many analytic tools are presented. The graphic methods serve several purposes. Construction of frequency plots or bar graphs and stem-and-leaf displays lets the investigator become familiar with the data and get a first impression of the batches of data to be compared. A stem-and-leaf display also is useful for detecting outliers and checking for data errors, and it is an efficient, practical way to view and report even large batches of data. Box plots, especially grouped box plots, allow immediate comparison of different sets of data; they are remarkably effective for comparing several batches at once. Inspection of such box plots is often all the investigator need do to conclude that several batches of habitat data are the same or different.

Four nonparametric techniques are given to statistically compare batches of habitat suitability curve data. Nonparametric tests are used because habitat data rarely meet the assumptions of parametric tests, and when the assumptions are met, these tests are nearly as powerful as their parametric counterparts. Nonparametric tests are often easier to compute and, more importantly, easier to understand, making them less likely to be misused.

Two of the tests, Kolmogorov-Smirnov and Mann-Whitney-Wilcoxon, are restricted to comparing two groups of data. The other two tests, Kruskal-Wallis and the multi-response permutation procedures, can be used to compare

two or more groups. All but the multi-response permutation procedures are restricted to analyzing one habitat variable at a time.

Other differences among the tests are also important. The Mann-Whitney-Wilcoxon and Kruskal-Wallis rank tests discern differences in the means or locations of the groups being compared. The other two tests are sensitive to differences in the scale, spread, shape, and location of the data. The Kolmogorov-Smirnov test does not require ranked data, but gives identical results for ranked or unranked data. The multi-response permutation procedures can be used on ranks or not, but give different results. They can also be constructed to mimic classic parametric and nonparametric tests and can be used on dichotomous categorical data.

REFERENCES

- Armour, C.L., R.J. Fisher, and J.W. Terrell. 1984. Comparison of the use of the habitat evaluation procedures (HEP) and the instream flow incremental methodology (IFIM) in aquatic analyses. U.S. Fish. Wildl. Serv. FWS/OBS-84/11. 30 pp.
- Armour, C.L., K.P. Burnham, and W.S. Platts. 1983. Field methods and statistical analyses for monitoring small salmonid streams. U.S. Fish. Wildl. Serv. FWS/OBS-83/33. 200 pp.
- Berry, K.J., K.L. Kvamme, and P.W. Mielke. 1983. Improvements in the permutation test for the spatial analysis of the distribution of artifacts into classes. *Amer. Antiquity* 48:547-553.
- Berry, K.J., and P.W. Mielke. 1983. Computation of finite population parameters and approximate probability values for multi-response permutation procedures (MRPP). *Commun. Statist. - Simulation Comput.* 12:83-107.
- Biondini, M.E., C.D. Bonham, and E.F. Redente. 1985. Secondary successional patterns in a sagebrush (*Artemisia tridentata*) community as they relate to soil disturbance and soil biological activity. *Vegetatio* 60:25-36.
- Bovee, K.D. 1986. Development and evaluation of habitat suitability criteria for use in the instream flow incremental methodology. Instream Flow Information Paper 21. U.S. Fish. Wildl. Serv. Biol. Rep. 86(7). 235 pp.
- Bovee, K.D., and T. Cochnauer. 1977. Development and evaluation of weighted criteria, probability-of-use curves for instream flow assessments: fisheries. Instream Flow Information Paper 3. U.S. Fish. Wildl. Serv. FWS/OBS-77/63. 38 pp.
- Chambers, J.M., W.S. Cleveland, B. Kleiner, and P.A. Tukey. 1983. Graphical methods for data analysis. Wadsworth International Group, Belmont, CA. 395 pp.
- Conover, W.J. 1980. Practical nonparametric statistics, 2nd ed. John Wiley and Sons, New York. 493 pp.
- Emerson, J.D., and D.C. Hoaglin. 1983. Stem-and-leaf displays. Pages 7-32 in D.C. Hoaglin, F. Mosteller, and J.W. Tukey, eds. Understanding robust and exploratory data analysis. John Wiley and Sons, Inc., New York.

- Emerson, J.D., and J. Strenio. 1983. Boxplots and batch comparison. Pages 58-93 in D.C. Hoaglin, F. Mosteller, and J.W. Tukey, eds. Understanding robust and exploratory data analysis. John Wiley and Sons, Inc., New York.
- Iman R.L., D. Quade, and D.A. Alexander. 1975. Exact probability levels for the Kruskal-Wallis test. Selected Tables in Mathematical Statistics 3:329-384.
- Kruskal, W.H., and W.A. Wallis. 1952. Use of ranks in one-criterion variance analysis. J. Am. Stat. Assoc. 47:583-621.
- McGill, R., J.W. Tukey, and W.A. Larsen. 1978. Variations of box plots. The American Statistician 32:12-16.
- Mielke, H.W., J.C. Anderson, K.J. Berry, P.W. Mielke, R.L. Chaney, and M. Leech. 1983. Lead concentrations in inner-city soils as a factor in the child lead problem. Amer. J. Public Health 73:1366-1369.
- Mielke, P.W. 1984. Meteorological applications of permutation techniques based on distance functions. Pages 813-830 in P.R. Krishnaiah and P.K. Sen, eds. Handbook of statistics, Vol. 4. North-Holland, Amsterdam.
- Mielke, P.W. 1986. Non-metric statistical analyses: some metric alternatives. Journal of Statistical Planning and Inference 13:377-387.
- Mielke, P.W., and K.J. Berry. 1982. An extended class of permutation techniques for matched pairs. Commun. Statist. - Theory Meth. A 11:1197-1207.
- Mielke, P.W., K.J. Berry, and E.S. Johnson. 1976. Multi-response permutation procedures for a priori classifications. Commun. Statist. - Theory Method A 5:1409-1424.
- Mielke, P.W., K.J. Berry, and G.W. Brier. 1981a. Application of multi-response permutation procedures for examining seasonal changes in monthly mean sea-level pressure patterns. Monthly Weather Rev. 109:120-126.
- Mielke, P.W., K.J. Berry, P.J. Brockwell, and J.S. Williams. 1981b. A class of nonparametric tests based on multiresponse permutation procedures. Biometrika 68:720-724.
- Mosteller, F., and R.E.K. Rourke. 1974. Sturdy statistics: nonparametrics and order statistics. Addison-Wesley, Reading, MA. 395 pp.
- Rohlf, F.J., and R.R. Sokal. 1981. Statistical tables, 2nd ed. W.H. Freeman and Company, New York. 219 pp.
- Sokal, R.R., and F.J. Rohlf. 1981. Biometry: the principles and practice of statistics in biological research, 2nd ed. W.H. Freeman and Company, San Francisco. 859 pp.

- Tukey, J.W. 1977. Exploratory data analysis. Addison-Wesley, Reading MA. 688 pp.
- Velleman, P.F., and D.C. Hoaglin. 1981. Applications, basics, and computing of exploratory data analysis. Duxbury Press, Boston, MA. 354 pp.
- Warren, W.G. 1986. On the presentation of statistical analysis: reason or ritual. Can. J. For. Res. 16:1185-1191.
- Zar, J.H. 1984. Biostatistical analysis, 2nd ed. Prentice-Hall, Inc., Englewood Cliffs, NJ. 718 pp.
- Zimmerman, G.M., H. Goetz, and P.W. Mielke. 1985. Use of an improved statistical method for group comparisons to study effects of prairie fire. Ecology 66:606-611.

REPORT DOCUMENTATION PAGE	1. REPORT NO. Biological Report 89(6)	2.	3. Recipient's Accession No.
4. Title and Subtitle Graphical and Statistical Procedures for Comparing Habitat Suitability Data			5. Report Date November 1988
7. Author(s) William L. Slauson			6.
9. Performing Organization Name and Address National Ecology Research Center U.S. Fish and Wildlife Service Creekside One Bldg., 2627 Redwing Rd. Fort Collins, CO 80526-2899			8. Performing Organization Rept. No.
12. Sponsoring Organization Name and Address Department of the Interior U.S. Fish and Wildlife Service Research and Development Washington, DC 20240			10. Project/Task/Work Unit No.
			11. Contract(C) or Grant(G) No. (C) (G)
15. Supplementary Notes			13. Type of Report & Period Covered
			14.
16. Abstract (Limit: 200 words) This paper presents a variety of graphic and statistical methods for exploring and comparing batches of data. The discussion and examples are directed towards habitat suitability curve data, but the methods also apply more generally. Data are explored, examined for errors or outliers, and informally compared by using frequency plots, bar graphs, stem-and-leaf displays, and grouped box plots. Four nonparametric techniques are given to statistically compare batches of habitat data. Two of the tests, Mann-Whitney-Wilcoxon and Kruskal-Wallis, use ranked data and are sensitive to differences in the location of batches. The two other tests, Kolmogorov-Smirnov and multi-response permutation procedures (MRPP), can be used on raw or ranked data and are sensitive to differences in scale, shape, and location. The MRPP can analyze more than one habitat variable at a time but, more importantly, allows statistical analyses commensurate with the geometry of the raw data. Nonparametric tests are used because biological data rarely meet the assumptions of parametric tests (e.g., normality), and when they do, nonparametric tests are nearly as powerful as their parametric counterparts. Nonparametric tests are often easier to compute and understand, making them less likely to be misused.			
17. Document Analysis a. Descriptors Nonparametric statistics Kolmogorov-Smirnov test b. Identifiers/Open-Ended Terms Multi-response permutation procedures (MRPP) Box plots Mann-Whitney-Wilcoxon test Habitat suitability data Kruskal-Wallis test, Stem-and-leaf displays c. COBATI Field/Group			
18. Availability Statement Release unlimited		19. Security Class (This Report) Unclassified	21. No. of Pages 58
		20. Security Class (This Page) Unclassified	22. Price



REGION 1

Regional Director
U.S. Fish and Wildlife Service
Lloyd Five Hundred Building, Suite 1692
500 N.E. Multnomah Street
Portland, Oregon 97232

REGION 2

Regional Director
U.S. Fish and Wildlife Service
P.O. Box 1306
Albuquerque, New Mexico 87103

REGION 3

Regional Director
U.S. Fish and Wildlife Service
Federal Building, Fort Snelling
Twin Cities, Minnesota 55111

REGION 4

Regional Director
U.S. Fish and Wildlife Service
Richard B. Russell Building
75 Spring Street, S.W.
Atlanta, Georgia 30303

REGION 5

Regional Director
U.S. Fish and Wildlife Service
One Gateway Center
Newton Corner, Massachusetts 02158

REGION 6

Regional Director
U.S. Fish and Wildlife Service
P.O. Box 25486
Denver Federal Center
Denver, Colorado 80225

REGION 7

Regional Director
U.S. Fish and Wildlife Service
1011 E. Tudor Road
Anchorage, Alaska 99503

